

# TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission

13

Application Number

10/652,653

Filing Date

August 28, 2003

First Named Inventor

TANAKA, Atsushi

Art Unit

2818

Examiner Name

Unassigned

Attorney Docket Number

16869S-092100US

PTO/SB/21 (09-04)

## ENCLOSURES (Check all that apply)



Fee Transmittal Form



Fee Attached



Amendment/Reply



After Final



Affidavits/declaration(s)



Extension of Time Request



Express Abandonment Request



Information Disclosure Statement



Certified Copy of Priority Document(s)



Reply to Missing Parts/ Incomplete Application



Reply to Missing Parts under 37 CFR 1.52 or 1.53



Drawing(s)



Licensing-related Papers



Petition to Make Special



Petition to Convert to a Provisional Application



Power of Attorney, Revocation Change of Correspondence Address



Terminal Disclaimer



Request for Refund



CD, Number of CD(s) \_\_\_\_\_

☐ Landscape Table on CD



After Allowance Communication to TC



Appeal Communication to Board of Appeals and Interferences



Appeal Communication to TC (Appeal Notice, Brief, Reply Brief)



Proprietary Information



Status Letter



Other Enclosure(s) (please identify below):



Return Postcard

Ten (10) cited references

Remarks

The Commissioner is authorized to charge any additional fees to Deposit Account 20-1430.

## SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name

Townsend and Townsend and Crew LLP

Signature

*Chun-Pok Leung*

Printed name

Chun-Pok Leung

Date

November 24, 2004

Reg. No.

41,405

## CERTIFICATE OF TRANSMISSION/MAILING

Express Mail Label: EV 530886551 US

I hereby certify that this correspondence is being deposited with the United States Postal Service with "Express Mail Post Office to Address" service under 37 CFR 1.10 on this date **November 24, 2004** and is addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below.

Signature

*Joy Salvador*

Typed or printed name

Joy Salvador

Date

November 24, 2004

# FEE TRANSMITTAL for FY 2005

Effective 10/01/2004. Patent fees are subject to annual revision.

Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT (\$) 130.00

Complete if Known

Application Number 10/652,653

Filing Date August 28, 2003

First Named Inventor TANAKA, Atsushi

Examiner Name Unassigned

Art Unit 2818

Attorney Docket No. 16869S-092100US

## METHOD OF PAYMENT (check all that apply)

☐ Check ☐ Credit Card ☐ Money Order ☐ Other ☐ None
☒ Deposit Account:Deposit  
Account  
Number

20-1430

Deposit  
Account  
Name

Townsend and Townsend and Crew LLP

The Director is authorized to: (check all that apply)

☒ Charge fee(s) indicated below ☒ Credit any overpayments☒ Charge any additional fee(s) or any underpayment of fee(s)☐ Charge fee(s) indicated below, except for the filing fee to the above-identified deposit account.

## FEE CALCULATION

## 1. BASIC FILING FEE

| Large Entity |          | Small Entity |          | Fee Description        | Fee Paid |
|--------------|----------|--------------|----------|------------------------|----------|
| Fee Code     | Fee (\$) | Fee Code     | Fee (\$) |                        |          |
| 1001         | 790      | 2001         | 395      | Utility filing fee     |          |
| 1002         | 350      | 2002         | 175      | Design filing fee      |          |
| 1003         | 550      | 2003         | 275      | Plant filing fee       |          |
| 1004         | 790      | 2004         | 395      | Reissue filing fee     |          |
| 1005         | 160      | 2005         | 80       | Provisional filing fee |          |

SUBTOTAL (1)

(\$0.00)

## 2. EXTRA CLAIM FEES FOR UTILITY AND REISSUE

|                    | Extra Claims | Fee from below | Fee Paid |
|--------------------|--------------|----------------|----------|
| Total Claims       | -** =        |                |          |
| Independent Claims | -** =        |                |          |
| Multiple Dependent |              |                |          |

| Large Entity |          | Small Entity |          | Fee Description  |
|--------------|----------|--------------|----------|--|
| Fee Code     | Fee (\$) | Fee Code     | Fee (\$) |  |
| 1202         | 18       | 2202         | 9        | Claims in excess of 20                                     |
| 1201         | 88       | 2201         | 44       | Independent claims in excess of 3                          |
| 1203         | 300      | 2203         | 150      | Multiple dependent claim, if not paid                      |
| 1204         | 88       | 2204         | 44       | ** Reissue independent claims over original patent         |
| 1205         | 18       | 2205         | 9        | ** Reissue claims in excess of 20 and over original patent |

SUBTOTAL (2)

(\$0.00)

\*\*or number previously paid, if greater; For Reissues, see above

## FEE CALCULATION (continued)

## 3. ADDITIONAL FEES

| Large Entity |          | Small Entity |          | Fee Description  | Fee Paid |
|--------------|----------|--------------|----------|--|----------|
| Fee Code     | Fee (\$) | Fee Code     | Fee (\$) |  |          |
| 1051         | 130      | 2051         | 65       | Surcharge - late filing fee or oath  |          |
| 1052         | 50       | 2052         | 25       | Surcharge - late provisional filing fee or cover sheet.                    |          |
| 1053         | 130      | 1053         | 130      | Non-English specification  |          |
| 1812         | 2,520    | 1812         | 2,520    | For filing a request for <i>ex parte</i> reexamination                     |          |
| 1804         | 920*     | 1804         | 920*     | Requesting publication of SIR prior to Examiner action                     |          |
| 1805         | 1,840*   | 1805         | 1,840*   | Requesting publication of SIR after Examiner action                        |          |
| 1251         | 110      | 2251         | 55       | Extension for reply within first month                                     |          |
| 1252         | 430      | 2252         | 215      | Extension for reply within second month                                    |          |
| 1253         | 980      | 2253         | 490      | Extension for reply within third month                                     |          |
| 1254         | 1,530    | 2254         | 765      | Extension for reply within fourth month                                    |          |
| 1255         | 2,080    | 2255         | 1,040    | Extension for reply within fifth month                                     |          |
| 1401         | 340      | 2401         | 170      | Notice of Appeal   |          |
| 1402         | 340      | 2402         | 170      | Filing a brief in support of an appeal                                     |          |
| 1403         | 300      | 2403         | 150      | Request for oral hearing   |          |
| 1451         | 1,510    | 1451         | 1,510    | Petition to institute a public use proceeding                              |          |
| 1452         | 110      | 2452         | 55       | Petition to revive - unavoidable   |          |
| 1453         | 1,330    | 2453         | 665      | Petition to revive - unintentional   |          |
| 1501         | 1,370    | 2501         | 685      | Utility issue fee (or reissue)   |          |
| 1502         | 490      | 2502         | 245      | Design issue fee   |          |
| 1503         | 660      | 2503         | 330      | Plant issue fee  |          |
| 1460         | 130      | 1460         | 130      | Petitions to the Commissioner  | 130      |
| 1807         | 50       | 1807         | 50       | Processing fee under 37 CFR 1.17(q)  |          |
| 1806         | 180      | 1806         | 180      | Submission of Information Disclosure Stmt                                  |          |
| 8021         | 40       | 8021         | 40       | Recording each patent assignment per property (times number of properties) |          |
| 1809         | 790      | 2809         | 395      | Filing a submission after final rejection (37 CFR § 1.129(a))              |          |
| 1810         | 790      | 2810         | 395      | For each additional invention to be examined (37 CFR § 1.129(b))           |          |
| 1801         | 790      | 2801         | 395      | Request for Continued Examination (RCE)                                    |          |
| 1802         | 900      | 1802         | 900      | Request for expedited examination of a design application                  |          |

Other fee (specify)

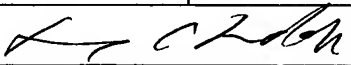
\*Reduced by Basic Filing Fee Paid

SUBTOTAL (3)

(\$130.00)

## SUBMITTED BY

Complete (if applicable)

|                   |   |                                   |        |           |                   |
|-------------------|---|-----------------------------------|--------|-----------|-------------------|
| Name (Print/Type) | Chun-Pok Leung  | Registration No. (Attorney/Agent) | 41,405 | Telephone | 650-326-2400      |
| Signature         |  |                                   |        | Date      | November 24, 2004 |

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.



PATENT  
Attorney Docket No.: 16869S-092100US  
Client Ref. No.: W1149-01EI

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re application of:

ATSUSHI TANAKA *et al.*

Application No.: 10/652,653

Filed: August 28, 2003

For: SAN/NAS INTEGRATED  
STORAGE SYSTEM

Customer No.: 20350

Examiner: Unassigned

Technology Center/Art Unit: 2818

Confirmation No.: 1913

**PETITION TO MAKE SPECIAL FOR  
NEW APPLICATION UNDER M.P.E.P.  
§ 708.02, VIII & 37 C.F.R. § 1.102(d)**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

This is a petition to make special the above-identified application under MPEP § 708.02, VIII & 37 C.F.R. § 1.102(d). The application has not received any examination by an Examiner.

(a) The Commissioner is authorized to charge the petition fee of \$130 under 37 C.F.R. § 1.17(i) and any other fees associated with this paper to Deposit Account 20-1430.

12/01/2004 BABRAHA1 00000053 201430 10652653  
01 FC:1464 130.00 DA

(b) All the claims are believed to be directed to a single invention. If the Office determines that all the claims presented are not obviously directed to a single invention, then Applicants will make an election without traverse as a prerequisite to the grant of special status.

(c) Pre-examination searches were made of U.S. issued patents, including a classification search, a computer database search, a keyword search, a literature search, and a foreign patent document search. The searches were performed on or around September 12, 2004, and were conducted by a professional search firm, Kramer & Amado, P.C. The classification search covered Class 711 (subclasses 113 and 114) for the U.S. and foreign subclasses identified above. The computer database search was conducted on the USPTO systems EAST and WEST. The keyword search was conducted in Class 707 (subclasses 10 and 100); Class 709 (subclasses 203 and 213); Class 711 (subclasses 112, 162, and 202); and Class 714 (subclasses 5, 6, 7, 11, and 13). The literature search was conducted on the Internet for relevant non-patent documents. The search for foreign patent documents was performed on the Espacenet and Delphion databases. The inventors further provided seven references considered most closely related to the subject matter of the present application (see references #5-10 below), which were cited in the Information Disclosure Statements filed on August 28, 2003 and September 12, 2003.

(d) The following references, copies of which are attached herewith, are deemed most closely related to the subject matter encompassed by the claims:

- (1) U.S. Patent Publication No. 2003/0105767 A1;
- (2) U.S. Patent Publication No. 2003/0204671 A1;
- (3) U.S. Patent No. 6,606,690 B2;
- (4) U.S. Patent Publication No. 2003/0023784 A1;
- (5) U.S. Patent Publication No. 2003/0135782 A1;
- (6) U.S. Patent Publication No. 2003/0023665 A1;
- (7) U.S. Patent Publication No. 2002/0152339 A1;

- (8) U.S. Patent Publication No. 2002/0116593 A1;
- (9) Japanese Patent Publication No. JP 2000-99281; and
- (10) Japanese Patent Publication No. JP 2002-14878.

(e) Set forth below is a detailed discussion of references which points out with particularity how the claimed subject matter is distinguishable over the references.

A. Claimed Embodiments of the Present Invention

The claimed embodiments relate to a system integrating a file system and a storage system.

Independent claim 1 recites a SAN/NAS integrated storage system comprising a plurality of disks for storing data; a plurality of controllers for receiving an I/O command from a host computer and controlling the disks in accordance with the I/O command; and a network for interconnecting the plurality of controllers. One of the plurality of controllers has a block interface as an interface to the host computer. Another of the plurality of controllers has a file interface as an interface to the host computer.

Independent claim 11 recites a storage system comprising a disk for storing data; a first adapter directly connected to a first network connected to a first host computer; a second adapter directly connected to the disk; and a second network directly connected to the first and second adapters for interconnecting the first and second adapters. The first adapter converts information in conformity with a first protocol received from the first host computer via the first network into information in conformity with a second protocol, and transferring the converted information to the second adapter via the second network.

Independent claim 25 recites a disk control system for transmitting and receiving a command and data to and from a host computer by utilizing an external protocol having a standard of at least one of "Ethernet II", "IEEE802.3" and "IEEE802.2" and accessing a recording disk by utilizing an internal protocol other than the external protocol. The disk control system comprises a file server unit and a channel adapter unit mounted on a single board or in a single housing. The file server unit comprises a server processor for controlling the file server unit and converting the command and the data from the external protocol into the internal protocol; a LAN controller for performing a communication control

of the command and the data by using the external protocol; a first internal bus for interconnecting the server processor and the LAN controller; and a host controller disposed between the server processor and the first internal bus for controlling transmission/reception of the command and the data and for controlling failure information. The channel adapter unit comprises a network data controller for controlling transmission/reception of the data to and from the recording disk; a control data controller for controlling transmission/reception of the command to and from the recording disk; a channel processor for controlling the channel adapter unit; and a second internal bus for interconnecting the network data controller and the control data controller. The disk control system further comprises a network data path connected between the first and second internal buses for transmission/reception of the data in accordance with the internal protocol, and a control data path for transmission/reception of the command in accordance with the internal protocol, respectively provided physically or logically independently between the file server unit and the channel adapter unit; and a management bus for transmission/reception of the failure information without involving the first and second internal buses.

Some of the benefits that may be derived are that the storage system has an interface of a file server, and that the command and the data are processed in parallel between a file server and the storage system. In addition, failure information of a file server is transferred via a path different from a path used for normal command and data, and the failure information is shared by the whole storage system. A fail-over of a file server with a failure is performed by using the failure information shared by the storage system.

B. Discussion of the References

None of the following references disclose a SAN/NAS integrated storage system comprising a plurality of disks for storing data; a plurality of controllers for receiving an I/O command from a host computer and controlling the disks in accordance with the I/O command; and a network for interconnecting the plurality of controllers, wherein one of the plurality of controllers has a block interface as an interface to the host computer and another of the plurality of controllers has a file interface as an interface to the host computer.

The references also do not teach a storage system comprising a disk for storing data; a first adapter directly connected to a first network connected to a first host computer; a

second adapter directly connected to the disk; and a second network directly connected to the first and second adapters for interconnecting the first and second adapters, wherein the first adapter converts information in conformity with a first protocol received from the first host computer via the first network into information in conformity with a second protocol, and transferring the converted information to the second adapter via the second network.

The references further fail to disclose a disk control system for transmitting and receiving a command and data to and from a host computer by utilizing an external protocol having a standard of at least one of "Ethernet II", "IEEE802.3" and "IEEE802.2" and accessing a recording disk by utilizing an internal protocol other than the external protocol; wherein the disk control system comprises a file server unit and a channel adapter unit mounted on a single board or in a single housing (the file server unit comprises a server processor, a LAN controller, and a first internal bus for interconnecting the server processor and the LAN controller, and a host controller disposed between the server processor and the first internal bus for controlling transmission/reception of the command and the data and for controlling failure information); and wherein the channel adapter unit comprises a network data controller, a control data controller, a channel processor, and a second internal bus for interconnecting the network data controller and the control data controller. Nor do the references disclose a network data path connected between the first and second internal buses for transmission/reception of the data in accordance with the internal protocol, and a control data path for transmission/reception of the command in accordance with the internal protocol, respectively provided physically or logically independently between the file server unit and the channel adapter unit; and a management bus for transmission/reception of the failure information without involving the first and second internal buses.

1. U.S. Patent Publication No. 2003/0105767 A1

This reference discloses a RAID storage system loaded with a NAS function and a SAN function, and a method of controlling the same. The storage system having the SAN and the NAS integrated with each other provides high availability, which makes possible the fail-over processing. The storage system includes a plurality of interfaces for the connection to the external network, a plurality of disks, and a shared memory to which the plurality of interfaces are accessible, wherein the plurality of interfaces are loaded with either one of block interfaces for executing an I/O request in disk blocks and file interfaces loaded

with file servers for executing an I/O request in files or both of these interfaces; and a file system to which a plurality of file servers are accessible in a sharing manner is constructed in a part of the plurality of disks. See [0020]-[0022] and [0037]-[0043].

2. U.S. Patent Publication No. 2003/0204671 A1

This reference discloses a storage system used in a computer system having interface controllers of a plurality of types having a block I/O interface controller having functions of a SAN and a file I/O interface controller having functions of a NAS. The disk controller (11) includes network channel adapters (1100), and Fibre Channel adapters (1110), disk adapters (1200), a shared memory (13), a cache memory (14), and a disk pool manager (15). The system includes a processor (11001), which controls overall operation of the adapter (1100). The LAN controller (11002) controls communication between the adapter and the LAN. The management networks controller (11003) controls communication between the adapter (1100) and the management network (16). See [0040]-[0052].

3. U.S. Patent No. 6,606,690 B2

This reference relates to a storage area network functionality in for Network Attached Storage that is attached to, and operating in a network. The SAN may be providing storage to hosts that communicate with the SAN to Small Computer Systems Interface (SCSI), Fibre Channel, and/or other data communication protocols on a first network. A storage appliance couples the SAN to the hosts. The system attaches the storage appliance to a second network, such that storage in the SAN may be accessed by hosts in the second network, such that storage in the SAN may be accessed by hosts in the second network as one or more NAS devices.

4. U.S. Patent Publication No. 2003/0023784 A1

This reference discloses a storage system having a plurality of disk controllers for accepting computer access through a SAN, and a plurality of file servers for accepting computer access through a LAN. The disk controllers and file servers are connected through a connection unit to a plurality of disk drive units in a disk pool. See [0019].

5. U.S. Patent Publication No. 2003/0135782 A1



This reference discloses a storage system including multiple slots for loading a block I/O interface controller, a file I/O interface controller, and any other kinds of interface controllers that are combined freely. The storage system includes a management table that manages fail-over-enabled devices by grouping those devices in accordance with the interface type and the domain to which each device belongs; and information table that directs a fail-over procedure; and fail-over controlling means that takes over the processing of a failed interface controller belonging to a fail-over-enabled group. The fail-over system offers several modalities for monitoring failures, selecting takeover controllers, and restoring functionality. The storage system solves conventional problems by providing a system that can mount a plurality of file systems, and that resists multiple failures detected in a fail-over server.

6. U.S. Patent Publication No. 2003/0023665 A1

This reference contains the same disclosure as reference #4 (2003/0023784).

7. U.S. Patent Publication No. 2002/0152339 A1

This reference discloses a storage system including a storage controller and storage media for reading data from or writing data to the storage media in response to SCSI, NFS, CIFS, or HTTP type read/write requests. The storage controller includes SCSI, NFS, CIFS, and HTTP interface adapters for receiving the read/write requests and effecting the reading of data to or the writing of data to the storage media.

8. U.S. Patent Publication No. 2002/0116593 A1

This reference discloses a system for responding to file system requests having file IDs V, a volume identifier specifying the file system being accessed, and R, an integer specifying the file within the file system being accessed. In order to allow a user having only an Ethernet port to connect to a storage system, a file server in the storage system and a host computer are connected by Ethernet and the file server and a channel adapter are connected by a block interface at a Fibre Channel port. The storage system has a file server which controls an IP interface for a user and is connected to the storage system by Fibre Channel via a mutual switch. With this approach, a file server process and a storage process are executed separately so that a system performance can be improved. Only one Fibre Channel cable is

used for a connection path to be used for transferring data or a command once. There is a high possibility that the performance is lowered if the load is heavy. Because only one Fibre Channel cable is used, failure information of a file server cannot be sent to the storage system. Thus, there is a high possibility that it takes a long time to complete a fail-over. Such a NAS system is associated with the problem of a lowered performance and a lowered reliability.

9. Japanese Patent Publication No. JP 2000-99281

This reference relates to a disk array control device (1) having a plurality of channel IF units (11), a plurality of disk IF units (12), a cache memory unit (14), and a shared memory unit (15), wherein the connection form between the plurality of IF units (11) and plurality of disk IF units (12) and the cache memory unit (15) is different from the connection form between the plurality of channel IF units (11) and plurality of disk IF units (12) and the shared memory unit (15). The most desirable mode of carrying out the invention is that the plurality of channel IF units (11) and plurality of disk IF units (12) are connected via a selector (13) to the cache memory unit (15) whereas the plurality of channel IF units (11) and plurality of disk IF units (12) are directly connected to the shared memory unit (15) with no selectors provided therebetween. The storage system does not have an IP interface to the host computer. A user without a file server cannot connect to the storage system. It is therefore necessary to prepare a new file server to connect the storage system. This results in an increase in the management cost and installation area.

10. Japanese Patent Publication No. JP 2002-14878

This reference discloses a server control bus (SMB) 4 arranged between an I/O processor 173 and host bridges 13 and 14. The environment setting processing of the host bridges 13 and 14 by the I/O processor 173 is operated through a server control bus (SMB) 4. The server control bus (SMB) 4 is constituted as an exclusive serial bus to be used for the maintenance control of this computer system so as to be made independent of a host bus 1 and PCI buses 2 and 3. Therefore, the server control bus 4 is used so that the host bridges 13 and 14 can be initialized under the control of the I/O processor 173 even when the failure of the main processor 11 is generated. The computer system can transmit failure information of a file server (main processor) directly to a disk control disk (I/O processor). However, actual

use of the failure information in the failure process is only after setting information of the file server is altered and the failure information is notified to an external management server via an IP network. Therefore, if a load of the IP network is heavy or there are some other reasons, the partner cannot acquire the failure information reliably or it takes time to transmit the failure information.

(f) In view of this petition, the Examiner is respectfully requested to issue a first Office Action at an early date.

Respectfully submitted,



Chun-Pok Leung  
Reg. No. 41,405

TOWNSEND and TOWNSEND and CREW LLP  
Two Embarcadero Center, 8<sup>th</sup> Floor  
San Francisco, California 94111-3834  
Tel: 650-326-2400  
Fax: 415-576-0300  
Attachments  
RL:rl  
60351287 v1

W 1149

**Disk array control device**

Patent Number: ☐ EP0987623  
Publication date: 2000-03-22  
Inventor(s): FUJIBAYASHI AKIRA (JP); FUJIMOTO KAZUHISA (JP); KANAI HIROKI (JP);  
MINOWA NOBUYUKI (JP); TANAKA ATSUSHI (JP)  
Applicant(s): HITACHI LTD (JP)  
Requested  
Patent: ☐ JP2000099281  
Application  
Number: EP19990114017 19990719  
Priority Number  
(s): JP19980264286 19980918  
IPC  
Classification: G06F3/06  
EC Classification: G06F3/06D, G06F13/12L  
Equivalents: ☐ US6385681  
Cited patent(s):

---

**Abstract**

---

A disk array control device (1) of the present invention has a plurality of channel IF units (11), a plurality of disk IF units (12), a cache memory unit (14), and a shared memory unit (15), wherein the connection form between said plurality of channel IF units (11) and plurality of disk IF units (12) and the cache memory unit (15) is different from the connection form between the plurality of channel IF units (11) and plurality of disk IF units (12) and the shared memory unit (15). The most desirable mode carrying out the invention is that the plurality of channel IF units (11) and plurality of disk IF units (12) are connected via a selector (13) to the cache memory unit (15) whereas the plurality of channel IF units (11) and plurality of disk IF units (12) are directly connected to the shared memory unit (15) with no selectors provided

therebetween.



---

Data supplied from the esp@cenet database - I2

特開 2000-99281

W 1149

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-99281

(P2000-99281A)

(43)公開日 平成12年4月7日(2000.4.7)

(51) Int.Cl.<sup>7</sup>

G O 6 F 3/06

識別記号

3 0 5

540

FI

G O B F 3/06

テーマコート\* (参考)

3 0 5 C      5 B 0 6 5

540

審査請求 未請求 請求項の数12 O L (全 19 頁)

(21)出願番号

特願平10-264286

(22) 出願日

平成10年9月18日(1998.9.18)

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 藤本 和久

東京都国分寺市東恋ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(72)発明者 田中 淳

東京都国分寺市東荻ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(74) 代理人 100068504

弁理士 小川 勝男

**最終頁に続く**

(54)【発明の名称】 ディスクアレイ制御装置

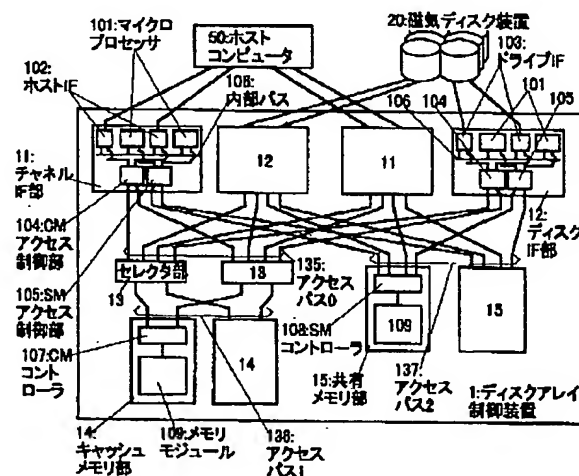
(57) 【要約】

【課題】 キャッシュメモリ及び共有メモリに格納されるデータの特性及びこれらのメモリへのアクセス特性を考慮した、スループットが高く、かつ、応答時間の短いディスクアレイ制御装置を提供することにある。

【解決手段】 上記課題は、複数のチャンネルＩＦ部と、複数のディスクＩＦ部と、キャッシュメモリ部と、共有メモリ部とを有し、前記複数のチャンネルＩＦ部及び前記複数のディスクＩＦ部と前記キャッシュメモリ部との間の接続形式が、前記複数のチャンネルＩＦ部及び前記複数のディスクＩＦ部と前記共有メモリ部との間の接続形式と異なることを特徴とするディスクアレイ制御装置により達成される。

【効果】 キャッシュメモリ及び共有メモリへのアクセスパスを増やしスループットを高くする一方、共有メモリへのアクセス時間を短くすることができる。

51



【特許請求の範囲】

【請求項1】 ホストコンピュータとのインターフェースを有する複数のチャンネルインターフェース部と、磁気ディスク装置とのインターフェースを有する複数のディスクインターフェース部と、前記磁気ディスク装置に対しリード／ライトされるデータを一時的に格納するキャッシュメモリ部と、チャンネルインターフェース部及びディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報を格納する共有メモリ部とを有し、各チャンネルインターフェース部は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、各ディスクインターフェース部は、前記磁気ディスク装置とのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行するディスクアレイ制御装置において、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間の接続形式が、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間の接続形式と異なることを特徴とするディスクアレイ制御装置。

【請求項2】 前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセクタ部を介して接続され、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間は、それぞれセクタ部を介せず直接接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項3】 前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセクタ部を介して接続され、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間はそれぞれは1対1接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項4】 前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセクタ部を介して接続され、前記複数のチャンネルインターフェース部、前記複数のディスクインターフェース部、及び前記共有メモリ部は共有バスに接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項5】 前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はスイッチを用いた相互結合網によって接続され、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間はそれぞれ直接接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項6】 前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はスイッチを用いた相互結合網によって接続され、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間はそれぞれ1対1接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項7】 ホストコンピュータとのインターフェースを有する複数のチャンネルインターフェース部と、磁気ディスク装置とのインターフェースを有する複数のディスクインターフェース部と、前記複数のチャンネルインターフェース部と前記複数のディスクインターフェース部とに接続され、前記磁気ディスク装置に対しリード／ライトされるデータを一時的に格納するキャッシュメモリ部と、前記複数のチャンネルインターフェース部と前記複数のディスクインターフェース部とに接続され、チャンネルインターフェース部及びディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報を格納する共有メモリ部とを有する制御装置において、

前記キャッシュメモリ部に接続されるアクセスパスの本数は、前記共有メモリ部に接続されるアクセスパスの本数より少ないことを特徴とするディスクアレイ制御装置。

【請求項8】 セクタ部をさらに有し、各チャンネルインターフェース部及び各ディスクインターフェース部と前記セクタ部とは、それぞれアクセスパスにより1対1に接続され、前記セクタ部と前記キャッシュメモリ部とは、アクセスパスにより接続され、各チャンネルインターフェース部及び各ディスクインターフェース部と前記セクタ部とを接続するアクセスパスの総数は、前記セクタ部と前記キャッシュメモリ部とを接続するアクセスパスの総数より多く、各チャンネルインターフェース部及び各ディスクインターフェース部と前記共有メモリ部との間はそれぞれアクセスパスにより1対1に接続されていることを特徴とする請求項7に記載のディスクアレイ制御装置。

【請求項9】 ホストコンピュータとの複数のインターフェース部と、磁気ディスク装置との複数のインターフェース部と、磁気ディスク装置のデータを一時的に格納する物理的に独立したキャッシュメモリ部と、制御情報を格納する物理的に独立した共有メモリ部とを有し、前記ホストコンピュータとの各インターフェース部及び前記磁気ディスク装置との各インターフェース部は、それぞれマイクロプロセッサと、前記キャッシュメモリ部へのアクセスを制御するキャッシュメモリ制御部と、前記共有メモリ部へのアクセスを制御する共有メモリアクセス制御部とを有するディスクアレイ制御装置において、各キャッシュメモリアクセス制御部と前記キャッシュメモリ部との間は、セクタ部を介してアクセスパスによ

り接続されており、各共有メモリアクセス制御部と前記共有メモリ部との間は、それぞれアクセスパスにより直接接続されていることを特徴とするディスクアレイ制御装置。

【請求項10】ホストコンピュータとの複数のインターフェース部と、磁気ディスク装置との複数のインターフェース部と、磁気ディスク装置のデータを一時的に格納する物理的に独立したキャッシュメモリ部と、制御情報を格納する物理的に独立した共有メモリ部とを有し、前記ホストコンピュータとの各インターフェース部及び前記磁気ディスク装置との各インターフェース部は、それぞれマイクロプロセッサと、前記キャッシュメモリ部へのアクセスを制御するキャッシュメモリ制御部と、前記共有メモリ部へのアクセスを制御する共有メモリアクセス制御部とを有するディスクアレイ制御装置において、各キャッシュメモリアクセス制御部と前記キャッシュメモリ部との間は、スイッチを用いた相互結合網によって接続されており、各共有メモリアクセス制御部と前記共有メモリ部との間は、それぞれアクセスパスにより直接接続されていることを特徴とするディスクアレイ制御装置。

【請求項11】各キャッシュメモリ制御部と前記キャッシュメモリ部との間のアクセスパスの帯域幅を、前記ホストコンピュータと前記ホストコンピュータとのインターフェース部との間の最大の全帯域幅の2倍以上としたことを特徴とする請求項9または請求項10の何れかに記載のディスクアレイ制御装置。

【請求項12】前記共有メモリ部及び前記キャッシュメモリ部を複数有し、該複数の共有メモリ部及び前記複数のキャッシュメモリ部で二重化されていることを特徴とする請求項1乃至請求項11の何れかに記載のディスクアレイ制御装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、データを複数の磁気ディスク装置に格納するディスクアレイ装置の制御装置に関する。

【0002】

【従来の技術】半導体記憶装置を記憶媒体とするコンピュータの主記憶のI/O性能に比べて、磁気ディスクを記憶媒体とするディスクサブシステム（以下「サブシステム」という。）のI/O性能は3～4桁程度小さく、従来からこの差を縮めること、すなわちサブシステムのI/O性能を向上させる努力がなされている。サブシステムのI/O性能を向上させるための1つの方法として、複数の磁気ディスク装置でサブシステムを構成し、データを複数の磁気ディスク装置に格納する、いわゆるディスクアレイと呼ばれるシステムが知られている。

【0003】図2は、従来のディスクアレイの構成を示す。ホストコンピュータ50とディスクアレイ制御装置

2との間のデータ転送を実行する複数のチャンネルIF部11と、磁気ディスク装置20とディスクアレイ制御装置2間のデータ転送を実行する複数のディスクIF部12と、磁気ディスク装置20のデータを一時的に格納するキャッシュメモリ部14と、ディスクアレイ制御装置2に関する制御情報（例えば、チャンネルIF部及びディスクIF部とキャッシュメモリ部14との間のデータ転送制御に関する情報）を格納する共有メモリ部15とを備え、キャッシュメモリ部14および共有メモリ部15は全てのチャンネルIF部11及びディスクIF部12からアクセス可能な構成となっている。このディスクアレイでは、チャンネルIF部11及びディスクIF部12と共有メモリ部15との間、及び、チャンネルIF部11及びディスクIF部12とキャッシュメモリ部14との間は1対1に接続される。以下、このような接続形態をスター接続と呼ぶ。

【0004】チャンネルIF部11は、ホストコンピュータ50と接続するためのインターフェース及びホストコンピュータ50に対する入出力を制御するマイクロプロセッサ（図示せず）を有している。また、ディスクIF部12は、磁気ディスク装置20と接続するためのインターフェース及び磁気ディスク装置20に対する入出力を制御するマイクロプロセッサ（図示せず）を有している。また、ディスクIF部12部は、RAID機能の実行も行う。

【0005】図3は、他の従来のディスクアレイの構成を示す。ホストコンピュータ50とディスクアレイ制御装置3間のデータ転送を実行する複数のチャンネルIF部11と、磁気ディスク装置20とディスクアレイ制御装置3間のデータ転送を実行する複数のディスクIF部12と、磁気ディスク装置20のデータを一時的に格納するキャッシュメモリ部14と、ディスクアレイ制御装置3に関する制御情報（例えば、チャンネルIF部及びディスクIF部とキャッシュメモリ部14との間のデータ転送制御に関する情報）を格納する共有メモリ部15を備え、各チャンネルIF部11及びディスクIF部12と共有メモリ部15間は共有バス130で接続され、各チャンネルIF部11部及びディスクIF部12とキャッシュメモリ部14間は共有バス131で接続される。以下、このような接続形態を共有バス接続形式と呼ぶ。

【0006】

【発明が解決しようとする課題】ディスクアレイのアーキテクチャーをスケーラブルなものとするには、ディスク制御装置に接続するディスク容量（論理ボリューム数）に応じ、ディスクIF部を増設し、また、必要なホストコンピュータとのチャンネル数に応じて、ディスクアレイ制御装置内のチャンネルIF部を増設する必要がある。しかし、第3図に示した共有バス接続形式のディスクアレイ制御装置では、一旦実装した共有バスの転送能力をチャンネルIF部、ディスクIF部の増設に応じて変

更することはできないので、チャンネルＩＦ部、ディスクＩＦ部の増設に柔軟に対応することが困難である。

【0007】また、第3図に示した共有バス接続形式のディスクアレイ制御装置では、ホストコンピュータとディスクアレイ制御装置との間のデータ転送を実行するチャンネルＩＦ部に設けられたホストコンピュータに対する入出力を制御するマイクロプロセッサ、及び磁気ディスク装置とディスクアレイ制御装置との間のデータ転送を実行するディスクＩＦ部に設けられた磁気ディスク装置に対する入出力を制御するマイクロプロセッサに高性能なプロセッサを使用した場合に、これらのプロセッサの性能に比べて、共有バスの転送能力がボトルネックになり、プロセッサの高速化に追従することが困難となる。

【0008】さらに、第3図に示した共有バス接続形式では、共有バスに接続された複数のチャンネルＩＦ部（または複数のディスクＩＦ部）の何れかのチャンネルＩＦ部（またはディスクＩＦ部）に障害が発生した場合に、障害の発生したチャンネルＩＦ部（またはディスクＩＦ部）を特定することが困難である。

【0009】一方、第2図に示したスター接続形式のディスクアレイ制御装置では、共有メモリ部またはキャッシュメモリ部に接続したアクセスバス数に比例して内部バス性能が増加させることができるので、チャンネルＩＦ部、ディスクＩＦ部の増設、または使用するプロセッサの性能に応じて、内部バス性能を増加させることが可能である。また、チャンネルＩＦ部及びディスクＩＦ部とキャッシュメモリ部との間、チャンネルＩＦ部及びディスクＩＦ部と共有メモリ部との間がスター接続されているため、障害の発生したチャンネルＩＦ部（またはディスクＩＦ部）を特定することも容易である。

【0010】スター接続形式のディスクアレイ制御装置では、搭載されるチャンネルＩＦ部またはディスクＩＦ部の数を増やした場合、チャンネルＩＦ部及びディスクＩＦ部とキャッシュメモリ部との間、及びチャンネルＩＦ部及びディスクＩＦ部と共有メモリとの間のアクセスバス数も増えることになる。また、ホストコンピュータとディスクアレイ制御装置との間の接続にファイバチャネル等の高速チャネルの採用等により、ディスクアレイ制御装置に要求されるスループットはさらに増大する方向にあり、このスループットの向上の要求を満たすためには、チャンネルＩＦ部及びディスクＩＦ部とキャッシュメモリ部との間、及びチャンネルＩＦ部及びディスクＩＦ部と共有メモリとの間のアクセスバス数を増やし、内部バス性能を向上させることが必要となる。

【0011】しかし、キャッシュメモリに格納される1つのデータのデータ量は、共有メモリに格納される1つの制御情報のデータ量よりもかなり大きい。一例を挙げれば、メインフレームに接続されるディスク制御装置では、キャッシュメモリに格納される1つのデータは数Kバイト程度（例えば2Kバイト）であるのに対し、共有

メモリに格納される1つの制御情報は数バイト程度（例えば4バイト）である。また、オープン系のホストコンピュータに接続されるディスク制御装置では、キャッシュメモリに格納される1つのデータは数十バイト程度

（例えば64バイト）であるのに対し、共有メモリに格納される1つの制御情報は数バイト程度（例えば4バイト）である。したがって、チャンネルＩＦ部及びディスクＩＦ部とキャッシュメモリ部との間で転送されるデータ量は、チャンネルＩＦ部及びディスクＩＦ部と共有メモリ部との間で転送されるデータ量に比べ、かなり多いので、チャンネルＩＦ部及びディスクＩＦ部とキャッシュメモリ部との間のアクセスバスのデータ幅は、チャンネルＩＦ部及びディスクＩＦ部と共有メモリ部との間のアクセスバスのデータ幅より広くとる必要がある。例えば、前者のアクセスバスは、16ビット幅のバスで構成され、後者は、4ビット幅のバスで構成される。そのため、チャンネルＩＦ部及びディスクＩＦ部とキャッシュメモリ部との間のアクセスバスの本数を増やすと、それらのアクセスバスを接続するキャッシュメモリ部のLSIのピン数が不足するという問題が生じる。また、ディスクアレイ制御装置のホストコンピュータへの応答時間を短くするためには、共有メモリ部に格納された制御情報へのアクセス時間をできるだけ短くすることも必要である。

【0012】そこで、本発明の目的は、キャッシュメモリ及び共有メモリに格納されるデータの特性及びこれらのメモリへのアクセス特性を考慮した、スループットの高く、ディスクアレイ制御装置、及びそれを用いたサブシステムを提供することにある。

【0013】より具体的には、本発明の目的は、チャンネルＩＦ部及びディスクＩＦ部とキャッシュメモリ部との間のアクセスバスはスループットが高く、チャンネルＩＦ部及びディスクＩＦ部と共有メモリとの間のアクセスバスはスループットが高く、かつアクセス時間が短いディスクアレイ制御装置、及びそれを用いたサブシステムを提供することにある。

【0014】

【課題を解決するための手段】上記目的は、ホストコンピュータとのインターフェースを有する複数のチャンネルインターフェース部と、磁気ディスク装置とのインターフェースを有する複数のディスクインターフェース部と、前記磁気ディスク装置に対しリード／ライトされるデータを一時的に格納するキャッシュメモリ部と、チャンネルインターフェース部及びディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報を格納する共有メモリ部とを有し、各チャンネルインターフェース部は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、各ディスクインターフェース部は、前記磁気ディスク装置とのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行するディスク



アレイ制御装置において、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間の接続形式が、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間の接続形式と異なることを特徴とするディスクアレイ制御装置により達成される。

【0015】好ましくは、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセクタ部を介して接続し、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間は、それぞれセクタ部を介せず直接接続する。

【0016】また、好ましくは、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセクタ部を介して接続し、前記複数のチャンネルインターフェース部、前記複数のディスクインターフェース部、及び前記共有メモリ部は共有バスに接続する。

【0017】また、好ましくは、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はスイッチを用いた相互結合網によって接続し、前記複数のチャンネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間はそれぞれ直接接続する。

【0018】その他、本願が開示する課題、及びその解決方法は、発明の実施形態の欄及び図面により明らかにされる。

#### 【0019】

【発明の実施の形態】以下、本発明の実施例を図面を用いて説明する。

【0020】[実施例1]図1に、本発明の一実施例を示す。

【0021】ディスクアレイ制御装置1は、ホストコンピュータ50との2つのインターフェース部（チャンネルIF部）11と、磁気ディスク装置20との2つのインターフェース部（ディスクIF部）12と、2つのセクタ部13と、2つのキャッシュメモリ部14と、2つの共有メモリ部15と、アクセスバス0:135と、アクセスバス1:136と、アクセスバス2:137とを有する。

【0022】チャンネルIF部11は、ホストコンピュータ50との2つのIF（ホストIF）102と、ホストコンピュータ50に対する入出力を制御する2つのマイクロプロセッサ101と、キャッシュメモリ部14へのアクセスを制御するアクセス制御部（CMアクセス制御部）104と、共有メモリ部15へのアクセスを制御するアクセス制御部（SMアクセス制御部）105とを有

し、ホストコンピュータ50とキャッシュメモリ部14間のデータ転送、及びマイクロプロセッサ101と共有メモリ部15間の制御情報の転送を実行する。マイクロプロセッサ101及びホストIF102は内部バス106によって接続され、CMアクセス制御部104は2つのホストIF102に直接接続されている。また、SMアクセス制御部105は2つのマイクロプロセッサ101に直接接続されている。

【0023】ディスクIF部12は、磁気ディスク装置20との2つのIF（ドライブIF）103と、磁気ディスク装置20に対する入出力を制御する2つのマイクロプロセッサ101と、キャッシュメモリ部14への1つのアクセス制御部（CMアクセス制御部）104と、共有メモリ部15への1つのアクセス制御部（SMアクセス制御部）105を有し、磁気ディスク装置20とキャッシュメモリ部14間のデータ転送、及びマイクロプロセッサ101と共有メモリ部15間の制御情報の転送を実行する。マイクロプロセッサ101及びドライブIF103は内部バス106によって接続され、CMアクセス制御部104は2つのドライブIF103に直接接続されている。また、SMアクセス制御部105は2つのマイクロプロセッサ101に直接接続されている。ディスクIF部はRAID機能の実行も行う。

【0024】キャッシュメモリ部14は、キャッシュメモリ（CM）コントローラ107とメモリモジュール109を有し、磁気ディスク装置20へ記録するデータを一時的に格納する。

【0025】共有メモリ部15は、共有メモリ（SM）コントローラ108とメモリモジュール109とを有し、ディスクアレイ制御装置1の制御情報（例えば、チャンネルIF部及びディスクIF部とキャッシュメモリ部1.4との間のデータ転送制御に関する情報）等を格納する。

【0026】CMアクセス制御部104には2本のアクセスバス0:135を接続し、それらを2つの異なるセクタ部13にそれぞれ接続する。セクタ部13には2本のアクセスバス1:136を接続し、それらを2つの異なるCMコントローラ107にそれぞれ接続する。したがってCMコントローラ107には、2つのセクタ部から1本ずつ、計2本のアクセスバス1:136が接続される。こうすることにより、1つのCMアクセス制御部104から1つのCMコントローラ107へのアクセスルートが2つとなる。これにより、1つのアクセスバスまたはセクタ部13に障害が発生した場合でも、もう1つのアクセスルートによりキャッシュメモリ部14へアクセスすることが可能となるため、耐障害性を向上させることができる。

【0027】SMアクセス制御部105には2本のアクセスバス2:137を接続し、そのアクセスバス2:137を2つの異なるSMコントローラにそれぞれ接続す

る。したがって、SMコントローラ108には、2つのチャンネルIF部11及び2つのディスクIF部12から1本ずつ、計4本のアクセスバス2:137が接続される。本実施例では、1つのSMアクセス制御部105と1つのSMコントローラ108の間には1本のアクセスバス2:137を接続したが、このアクセスバス2:137を2本に増やすことにより、1つのSMアクセス制御部105から1つのSMコントローラ108へのアクセスルートが2つとなるため、耐障害性を向上させることができる。

【0028】セクタ部13には、2つのチャンネルIF部11と、2つのディスクIF部12からそれぞれ1本ずつ、計4本のアクセスバス0:135が接続される。また、セクタ部13には、2つのキャッシュメモリ部14へのアクセスバス1:136が1本ずつ、計2本接続される。

【0029】アクセスバス0:135とアクセスバス1:136の間に上記のようなバス数の関係があるため、セクタ部13ではチャンネルIF部11及びディスクIF部12からの4本のアクセスバス0:135からの要求の内、キャッシュメモリ部14へのアクセスバス1:136の数に相当する2個だけを選択して実行する機能を持つ。

【0030】本実施例の大きな特徴は、チャンネルIF部11及びディスクIF部12とキャッシュメモリ部14との間の接続形式と、チャンネルIF部11及びディスクIF部12と共有メモリ部15との間の接続形式とが異なる点にある。このような構成にした理由を図1及び図2を用いて説明する。図2に示したスター接続形式のディスクアレイ制御装置2では、ディスクアレイ制御装置2に搭載されるチャンネルIF部11またはディスクIF部12の数を増やした場合、チャンネルIF部11及びディスクIF部12とキャッシュメモリ部14との間、及びチャンネルIF部11及びディスクIF部12と共有メモリ部15との間のアクセスバス数も増えることになる。また、スター接続形式のディスクアレイ制御装置2において、スループットを向上させるためには、チャンネルIF部11及びディスクIF部12とキャッシュメモリ部14との間、及びチャンネルIF部11及びディスクIF部12と共有メモリ部15との間のアクセスバス数を増やし、内部バス性能を向上させることが有効である。

【0031】しかし、キャッシュメモリ部14に格納される1つのデータのデータ量は、共有メモリ部15に格納される1つの制御情報のデータ量よりもかなり大きい。一例を挙げれば、メインフレームに接続されるディスク制御装置では、キャッシュメモリ部14に格納される1つのデータは数Kバイト程度（例えば2Kバイト）であるのに対し、共有メモリ部15に格納される1つの制御情報は数バイト程度（例えば4バイト）である。ま

た、オープン系のホストコンピュータに接続されるディスク制御装置では、キャッシュメモリ部14に格納される1つのデータは数十バイト程度（例えば64バイト）であるのに対し、共有メモリ部15に格納される1つの制御情報は数バイト程度（例えば4バイト）である。したがって、チャンネルIF部11及びディスクIF部12とキャッシュメモリ部14との間で転送されるデータ量は、チャンネルIF部11及びディスクIF部12と共有メモリ部15との間で転送されるデータ量に比べかなり多いので、チャンネルIF部11及びディスクIF部12とキャッシュメモリ部14との間のアクセスバスのデータ幅は、チャンネルIF部11及びディスクIF部12と共有メモリ部15との間のアクセスバスのデータ幅より広くとる必要がある。例えば、前者のアクセスバスは16ビット幅のバスで構成され、後者は、8ビット幅のバスで構成される。そのため、チャンネルIF部11及びディスクIF部12とキャッシュメモリ部14との間のアクセスバス本数を増やすと、キャッシュメモリ部14内のキャッシュメモリコントローラ（図2ではキャッシュメモリコントローラを図示していない）のLSIのピン数不足、またはキャッシュメモリ部14を実装するパッケージにおいてコネクタのピン数不足という問題が生じる。そこで、本実施例では、図1に示すように、チャンネルIF部11及びディスクIF部12とキャッシュメモリ部14との間をセクタ部13を介して接続することにより、キャッシュメモリ部14に直接接続されるアクセスバス数を削減している。

【0032】一方、上述したように、共有メモリ部13へ格納する1つの制御情報のデータ長はキャッシュメモリ部14に格納する1つのデータのデータ長に比べかなり小さいので、チャンネルIF部11及びディスクIF部12と共有メモリ部15との間のアクセスバスのデータ幅は、チャンネルIF部11及びディスクIF部12とキャッシュメモリ部14との間のアクセスバスのデータ幅の半分以下とすることが可能である。したがって、共有メモリ部15へのアクセスバス数を増やしても共有メモリ部内の共有メモリメモリコントローラ（図2では共有メモリコントローラを図示していない）のLSIのピン数不足等の問題が生じることは少ない。

【0033】また、ディスクアレイ制御装置1のホストコンピュータ50への応答時間を短くするためには、共有メモリ部15に格納される制御情報へのアクセス時間をできるだけ短くする必要もある。しかし、図1に示したCMアクセス制御部104とCMコントローラ107間のように、SMアクセス制御部105とSMコントローラ108との間をセクタ部を介して接続すると、セクタ部での処理のオーバーヘッドにより、共有メモリ部15に格納される制御情報へのアクセス時間を短くすることができない。

【0034】そこで、本実施例では、チャンネルIF部1

1及びディスクIF部12と共有メモリ部15との間をセクタ部を介さず直接接続することにより、チャンネルIF部11及びディスクIF部12と共有メモリ部15間に複数のアクセスパスを設け、スループットを向上させる一方、セクタ部での処理オーバーヘッドをなくし、チャンネルIF部11、及びディスクIF部12から共有メモリ部15へのアクセス時間を短縮にしている。

【0035】なお、本実施例では、耐障害性の向上という観点から、セクタ部13、キャッシュメモリ部14、及び共有メモリ部15をそれぞれ二重化しているが、これらを二重化しなくても、上述の効果をえられることは言うまでもない。

【0036】図4は、CMアクセス制御部104内の構成を示している。CMアクセス制御部104は、セクタ302と、アドレス、コマンド、データを一時格納するバケットバッファ303と、セクタ部13に繋がるアクセスパス0:135とのバスIF301と、データのエラーチェック部300と、データ転送制御部310を有する。セクタ302の2つのポートはデータ線210でホストIF102あるいはドライブIF103に接続される。また、セクタ302の他の2つのポートはバスIF301に接続される。バスIF301はアクセスパス0:135でセクタ部13に接続される。データ転送制御部310は、制御線1:211でホストIF102あるいはドライブIF103に接続され、制御線2:212でセクタ部13内のデータ転送制御部315に接続される。また、データ転送制御部310は、アービタ308によりホストIF102あるいはドライブIF103からのアクセス要求のアービトレーションを行い、セクタ302の切り替えを行う。

【0037】図6は、セクタ部13内の構成を示している。セクタ部13は、チャンネルIF部11及びディスクIF部12に繋がるアクセスパス0:135との4つのバスIF301と、CMコントローラ107に繋がるアクセスパス1:136との2つのバスIF301と、両者間を互いに接続するセクタ306と、バケットバッファ303と、データのエラーチェック部300と、CMアクセス制御部104から送出されたアドレス及びコマンドを解析するアドレス・コマンド(a dr、c m d)解析部305と、データ転送制御部315を有する。データ転送制御部315は、制御線2:212でCMアクセス制御部104内のデータ転送制御部310に接続され、制御線3:213でCMコントローラ107内のデータ転送制御部315に接続される。また、データ転送制御部315は、アービタ308により、a d r、c m d解析部305で解析した4本のアクセスパス0:135からのアクセス要求のアービトレーションを行い、セクタ306の切り替えを行う。バケットバッファ303は、アクセスパス0:135側のバスとアクセスパス1:136側のバスでデータ転送速度に差があ

る場合、速度差を吸収するために、転送するデータの一部または全部をバッファリングする。

【0038】a d r、c m d解析部305は、アドレス及びコマンドを格納するバッファと、a d r抽出部と、c m d抽出部を有する(図示していない)。a d r、c m d解析部305では、CMアクセス制御部104に接続される4本のアクセスパス0:135それぞれに1つずつ割り当てられたバッファに、アドレス、コマンドを格納する。a d r抽出部及びc m d抽出部では、アクセスするCMコントローラ107とアクセスの種類を割り出し、データ転送制御部315内のアービタ308へ送出する。

【0039】図7は、キャッシュメモリ部14内の構成を示している。キャッシュメモリ部14は、CMコントローラ107とメモリモジュール109を有する。CMコントローラ107は、セクタ部13に繋がるアクセスパス1:136との2つのバスIF301と、セクタ304と、データを一時格納するバケットバッファ303と、データのエラーチェック部300と、メモリモジュール109へのアクセスを制御するメモリ制御部307と、CMアクセス制御部104から送出されたアドレス及びコマンドを解析するa d r、c m d解析部305と、データ転送制御部315を有する。データ転送制御部315は、制御線3:213でセクタ部13内のデータ転送制御部315に接続される。また、データ転送制御部315は、アービタ308により、a d r、c m d解析部305で解析した2本のアクセスパス1:136からのアクセス要求のアービトレーションを行い、セクタ304の切り替えを行う。

【0040】a d r、c m d解析部305は、バッファと、a d r抽出部と、c m d抽出部を有する(図示していない)。a d r、c m d解析部305では、CMコントローラ107に接続される2本のアクセスパス1:136それぞれに1つずつ割り当てられたバッファに、アドレス、コマンドを格納する。a d r抽出部及びc m d抽出部では、アクセスするメモリのアドレスとアクセスの種類を割り出し、メモリ制御部307へ送出する。また、2本のアクセスパス1:136からのアクセス要求をデータ転送制御部315内のアービタ308へ送出する。

【0041】次に、キャッシュメモリ部14へのアクセス時の手順について述べる。キャッシュメモリ部14へアクセスする場合、マイクロプロセッサ101は、ホストIF102あるいはドライブIF103へ、キャッシュメモリ部14へのアクセス開始を指示する。

【0042】アクセス開始の指示を受けたホストIF102あるいはドライブIF103は、制御線1:211によりCMアクセス制御部104内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとともに、データ線210を通してアドレス、コマンド、デ

ータ（データの書き込み時のみ）を送出する。

【0043】CMアクセス制御部104は、データ線210を通して送られてきたアドレス、コマンド、データ（データの書き込み時のみ）をバッファ303に格納する。データ転送制御部310はアービトレーションを行ってバスIF301の使用権を決定し、セクタ302を切り替える。

【0044】図9は、キャッシュメモリ部14へデータを書き込む場合の、CMアクセス制御部104からCMコントローラ107へのアクセスの流れを示している。CMアクセス制御部104内のデータ転送制御部310は、アービトレーションによってアクセスバス0:135の使用権が決定されると、制御線2:212によってセクタ部13内のデータ転送制御部315へアクセス開始を示す信号（REQ）を出す（ステップ501）。続いて、アドレス及びコマンドを送出する（ステップ502）。

【0045】セクタ部13内のデータ転送制御部315は、CMアクセス制御部104からREQ信号を受け取ると、次にアクセスバス0:135を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行う（ステップ503）。アービトレーションの結果、アクセスバス1:136への接続権を得たら、データ転送制御部315はセクタ306を切り替える（ステップ504）とともに、制御線2:212により、CMアクセス制御部104内のデータ転送制御部310へ、アクセスバス1:136への接続権が得られたことを示す信号（ACK）を返す（ステップ505）。次にデータ転送制御部315は、制御線3:213によってCMコントローラ107内のデータ転送制御部315へアクセス開始を示す信号（REQ）を出す（ステップ506）。続いて、アドレス及びコマンドを送出する（ステップ507）。

【0046】CMアクセス制御部104はACK信号を受けると、バッファ303からデータを読み出し、セクタ302、バスIF301を介してアクセスバス0:135へ送出する。セクタ部13は、アクセスバス0:135を通して送られてきたデータを、バスIF301及びセクタ306を介してアクセスバス1:136へ送出する（ステップ509）。

【0047】CMコントローラ107内のデータ転送制御部315は、制御線3:213によってREQ信号を受け取ると、次にアクセスバス1:136を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行い（ステップ508）、セクタ304を切り替える。アクセスバス1:136を通して送られてくるデータはバッファ303に格納する。アービトレーションの結果、メモリモジュール10

9へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う（ステップ510）。次に、バッファ303からデータを読み出し、セクタ304を介してメモリモジュール109へ書き込む（ステップ511）。

【0048】メモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス（STATUS）を生成する（ステップ512）。次に、ステータスをセクタ部13を介してCMアクセス制御部104へ送出する（ステップ513）。セクタ部13内のデータ転送制御部315はステータスを受け取ると、CMコントローラ107へのREQ信号をオフする（ステップ514）。CMアクセス制御部104内のデータ転送制御部310はステータスを受け取ると、セクタ部13へのREQ信号をオフする（ステップ515）。セクタ部13内のデータ転送制御部315はCMアクセス制御部104からのREQ信号のオフを確認すると、CMアクセス制御部104へのACK信号をオフする（ステップ516）。

【0049】CMアクセス制御部104内のデータ転送制御部310はステータスを受け取ると、制御線1:211により、バスIF102あるいはドライブIF103へキャッシュメモリ部14へのアクセスの終了を報告する。

【0050】キャッシュメモリ部14からデータを読み出す場合の、CMアクセス制御部104からCMコントローラ107へのアクセスの流れは、ステップ501から508までとステップ512以降は、データの書き込みの場合と同じである。

【0051】ここでCMアクセス制御部104は、ステップ505でACK信号を受けると、データの受信待ち状態に入る。

【0052】ステップ508でメモリアクセス権を得ると、CMコントローラ107はメモリモジュール109からデータを読み出し、セクタ304、バスIF301を介してアクセスバス1:136にデータを送出する。

【0053】セクタ部13は、アクセスバス1:136を通してデータを受信すると、バスIF301及びセクタ306を介してアクセスバス0:135にデータを送出する。

【0054】CMアクセス制御部104は、アクセスバス0:135を通してデータを受信すると、セクタ302、データ線210を介してバスIF102あるいはドライブIF103へデータを送出する。

【0055】図5は、SMアクセス制御部105内の構成を示している。SMアクセス制御部104は、セクタ302と、アドレス、コマンド、データを一時格納するバッファ303と、SMコントローラ108

に繋がるアクセスバス2:137とのバスIF301と、データのエラーチェック部300と、データ転送制御部310を有する。セクタ302の2つのポートはデータ線220でマイクロプロセッサ101に接続される。また、セクタ302の他の2つのポートはバスIF301に接続される。バスIF301はアクセスバス2:137でSMコントローラ108に接続される。データ転送制御部310は、制御線5:221でマイクロプロセッサ101に接続され、制御線6:222でSMコントローラ108内のデータ転送制御部315に接続される。また、データ転送制御部310は、アービタ308によりマイクロプロセッサ101からのアクセス要求のアービトレーションを行い、セクタ302の切り替えを行う。

【0056】図8は、共有メモリ部15内の構成を示している。共有メモリ部15は、SMコントローラ108とメモリモジュール109を有する。SMコントローラ108は、SMアクセス制御部105に繋がるアクセスバス2:137との4つのバスIF301と、セクタ309と、データを一時格納するパケットバッファ303と、データのエラーチェック部300と、メモリモジュール109へのアクセスを制御するメモリ制御部307と、SMアクセス制御部105から送出されたアドレス及びコマンドを解析するadr、cmd解析部305と、データ転送制御部315を有する。データ転送制御部315は、制御線6:222でSMアクセス制御部105内のデータ転送制御部310に接続される。また、データ転送制御部315は、アービタ308により、adr、cmd解析部305で解析した4本のアクセスバス2:137からのアクセス要求のアービトレーションを行い、セクタ309の切り替えを行う。

【0057】adr、cmd解析部305は、バッファと、adr抽出部と、cmd抽出部を有する(図示していない)。adr、cmd解析部305では、SMコントローラ108に接続される4本のアクセスバス2:137それぞれに1つずつ割り当てられたバッファに、アドレス、コマンドを格納する。adr抽出部及びcmd抽出部では、アクセスするメモリのアドレスとアクセスの種類を割り出し、メモリ制御部307へ送出する。また、4本のアクセスバス2:137からのアクセス要求をデータ転送制御部315内のアービタ308へ送出する。

【0058】次に、共有メモリ部15へのアクセス時の手順について述べる。共有メモリ部15へアクセスする場合、マイクロプロセッサ101は、制御線5:221によりSMアクセス制御部105内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとともに、データ線220を通してアドレス、コマンド、データ(データの書き込み時のみ)を送出する。

【0059】SMアクセス制御部105は、データ線2

20を通して送られてきたアドレス、コマンド、データ(データの書き込み時のみ)をパケットバッファ303に格納する。データ転送制御部310はアービトレーションを行ってバスIF301の使用権を決定し、セクタ302を切り替える。

【0060】図10は、共有メモリ部15へデータを書き込む場合の、SMアクセス制御部105からSMコントローラ108へのアクセスの流れを示している。SMアクセス制御部105内のデータ転送制御部310は、アービトレーションによってアクセスバス2:137の使用権が決定されると、制御線6:222によってSMコントローラ108へアクセス開始を示す信号(REQ)を出す(ステップ601)。続いて、アドレス、コマンド、及びデータを連続して送出する(ステップ602)。

【0061】SMコントローラ108内のデータ転送制御部315は、制御線6:222によってREQ信号を受け取ると、次にアクセスバス2:137を通して送られてくるアドレス、コマンド、及びデータを受信する。アドレスとコマンドは、adr、cmd解析部305で解析し、アクセス要求に基づいてアービトレーションを行い(ステップ603)、セクタ309を切り替える。データはパケットバッファ303に格納する。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ604)。次に、パケットバッファ303からデータを読み出し、セクタ309を介してメモリモジュール109へ書き込む(ステップ605)。

【0062】メモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ606)。次に、ステータスをSMアクセス制御部105へ送出する(ステップ607)。SMアクセス制御部105内のデータ転送制御部310はステータスを受け取ると、SMコントローラ108へのREQ信号をオフする(ステップ608)。

【0063】SMアクセス制御部105内のデータ転送制御部310はステータスを受け取ると、制御線5:221により、マイクロプロセッサ101へ共有メモリ部15へのアクセスの終了を報告する。

【0064】共有メモリ部15からデータを読み出す場合のSMアクセス制御部105からSMコントローラ108へのアクセスの流れは、ステップ601から604までとステップ606以降は、データの書き込みの場合と同じである。

【0065】ステップ604でメモリアクセスの前処理を行った後、SMコントローラ108はメモリモジュール109からデータを読み出し、セクタ309、バス

IF301を介してアクセスバス2:137にデータを送出する。

【0066】SMアクセス制御部105は、アクセスバス2:137を通してデータを受信すると、セクタ302、データ線220を介してマイクロプロセッサ101へデータを送出する。

【0067】本実施例のディスクアレイ制御装置1では、ホストコンピュータ50とのチャネルを2つ有するチャネルIF部12を複数搭載しており、それらのチャネルをそれぞれ異なるホストコンピュータに接続することが可能である。そうした場合、ホストコンピュータに接続した各チャネルからの要求全てを並列に処理する必要がある。

【0068】ところでディスクアレイ制御装置1では、ホストコンピュータ50へデータを読み出す場合、磁気ディスク装置20に格納されたデータをディスクIF部12を介してキャッシュメモリ部14へ書き込み、そのデータをキャッシュメモリ部14から読み出して、チャネルIF部11を介してホストコンピュータ50へ送る。またホストコンピュータ50からディスクアレイ制御装置1へデータを書き込む場合は、ホストコンピュータ50からチャネルIF部12へ送られてきたデータをキャッシュメモリ部14へ書き込み、そのデータをキャッシュメモリ部14から読み出して、ディスクIF部12を介して磁気ディスク装置20に書き込む。さらにデータのパリティを生成して磁気ディスク装置20に書き込むため、ディスクIF部12とキャッシュメモリ部14間でさらに2~3回のアクセスが行われる。

【0069】したがって、ホストコンピュータに接続した各チャネルからの要求全てを並列に処理するためには、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14間のスループットをホストコンピュータ50とチャネルIF部11間の最大のスループットの2倍以上にしなければならない。

【0070】本実施例では、CMアクセス制御部104とCMコントローラ107間のアクセスバスの帯域幅、及びCMコントローラ107とメモリモジュール109間の全帯域幅を、チャネルIF部11とホストコンピュータ50間の最大の帯域幅の2倍以上に設定する。これにより、全チャネルIF部11を並列に動作させることが可能となる。

【0071】本実施例によれば、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14間のスループットを高くすること、チャネルIF部11及びディスクIF部12と共有メモリ部15間のスループットを高く、且つアクセス時間を短くすることの両方が可能となる。これによって、スループットが高く、且つ応答時間の短いディスクアレイ制御装置を提供できる。

【0072】ここで、図15に示すように、チャネルIF部11及びディスクIF部12とキャッシュメモリ部

14間をスイッチ(SW)16を用いた相互結合網140で接続する。この場合にも、図1に示したセクタ部13を介して接続した構成と同様に、キャッシュメモリ部14へ複数のアクセスバスを設けることができるため、スループットを高めることが可能となる。

【0073】また図16に示すように、1つのCMアクセス制御部104へ接続されるアクセスバス0:135の本数を図1の構成の倍の4本に増やしたディスクアレイ制御装置1においても、本実施例を実施する上で問題は無い。ホストIF及びドライブIFとして、今後はファイバチャネル等のスループットが100MB/s以上の高速IFが使用されることが多くなると考えられる。チャネルIF部11及びディスクIF部12でのスループットのバランスを考えると、1つのCMアクセス制御部104に繋がる全アクセスバス0:135のスループットは、チャネルIF部11内の全ホストIF102、またはディスクIF部12内の全ドライブIF103のスループットと同等以上にしなければならない。上記のようにファイバチャネル等の高速IFを使用する場合は、図16に示すように、1つのCMアクセス制御部104に繋がるアクセスバス0:135の本数を増やすことで、アクセスバス0:135のスループットをホストIF102あるいはドライブIF103のスループット以上にすることができる。

【0074】なお、図15及び図16のディスクアレイ制御装置では、耐障害性の向上という観点から、セクタ部13、キャッシュメモリ部14、及び共有メモリ部15をそれぞれ二重化しているが、これらを二重化しなくても、上述の効果をえられることは言うまでもない。

【0075】〔実施例2〕図1において、2つのキャッシュメモリ部14間で、メモリ領域の全部または、一部を二重化し、キャッシュメモリ部14へのデータの書き込み時に二重化した2つの領域に同じデータを書き込むことにより、データの信頼性を上げることが可能となる。

【0076】二重化した2つのキャッシュメモリ部14へデータを書き込む場合の手順は以下になる。

【0077】アクセス開始の指示を受けたホストIF102あるいはドライブIF103は、制御線1:211によりCMアクセス制御部104内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとともに、データ線210を通して2つのアドレス、コマンドと、1つのデータを送出する。

【0078】CMアクセス制御部104は、データ線210を通して送られてきた2つのアドレス、コマンドと、1つのデータをパケットバッファ303に格納する。データ転送制御部310はアービトレーションを行ってバスIF301の使用権を決定し、セクタ302を切り替える。

【0079】図11は、二重化した2つのキャッシュメ

メモリ部14へデータを書き込む場合の、CMアクセス制御部104から2つのCMコントローラa、b：107へのアクセスの流れを示している。CMアクセス制御部104内のデータ転送制御部310は、アービトレーションによってアクセスバス0：135の使用権が決定されると、制御線2：212によってセクタ部13内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ701)。続いて、アドレス及びコマンドを2つ連続して送出する(ステップ702)。

【0080】セクタ部13内のデータ転送制御部315は、CMアクセス制御部104からREQ信号を受け取ると、次にアクセスバス0：135を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行う(ステップ703)。アービトレーションの結果、CMコントローラa、b：107への2本のアクセスバス1：136への接続権の両方を得たら、データ転送制御部315はセクタ306を切り替える(ステップ704)とともに、制御線2：212により、CMアクセス制御部104内のデータ転送制御部310へ、CMコントローラa、b：107への接続権が得られたことを示す信号(ACK)を返す(ステップ705)。次にデータ転送制御部315は、制御線3：213によって2つのCMコントローラ107内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ706a、706b)。続いて、CMコントローラa、b：107へアドレス及びコマンドを1つずつ送出する(ステップ707a、707b)。

【0081】CMアクセス制御部104はACK信号を受けると、バケットバッファ303からデータを読み出し、セクタ302、バスIF301を介してアクセスバス0：135へ送出する(ステップ709)。セクタ部13は、アクセスバス0：135を通して送られてきた1つのデータを、バスIF301及びセクタ306を介して2つのアクセスバス1：136の両方へ送出する(ステップ709a、709b)。

【0082】CMコントローラa、b：107内のデータ転送制御部315は、制御線3：213によってREQ信号を受け取ると、次にアクセスバス1：136を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行い(ステップ708a、708b)、セクタ304を切り替える。アクセスバス1：136を通して送られてくるデータはバケットバッファ303に格納する。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ710a、710b)。次に、バケットバッファ303からデータを読み

出し、セクタ304を介してメモリモジュール109へ書き込む(ステップ711a、711b)。

【0083】メモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ712a、712b)。次に、ステータスをセクタ部13を介してCMアクセス制御部104へ送出する(ステップ713a、713b)。セクタ部13内のデータ転送制御部315はステータスを受け取ると、CMコントローラa、b：107へのREQ信号をそれぞれオフする(ステップ714a、714b)。また、セクタ部13はCMコントローラa、b：107の両方からステータスを受け取ったら、それらを続けてCMアクセス制御部へ送出する(ステップ713)。CMアクセス制御部104内のデータ転送制御部310は2つのステータスを受け取ると、セクタ部13へのREQ信号をオフする(ステップ715)。セクタ部13内のデータ転送制御部315はCMアクセス制御部104からのREQ信号のオフを確認すると、CMアクセス制御部104へのACK信号をオフする(ステップ716)。

【0084】CMアクセス制御部104内のデータ転送制御部310はステータスを受け取ると、制御線1：211により、ホストIF102あるいはドライブIF103へキャッシュメモリ部14へのアクセスの終了を報告する。

【0085】上記のように二重化したキャッシュメモリ部14にデータを二重に書き込む場合、二重に書き込むデータの一方のデータの書き込み処理を行っている間に、もう一方のデータが他のアクセスバスからの書き込み要求によって書き換えられるのを防ぐ必要がある。本実施例では、共有メモリ部15にキャッシュメモリ部14のディレクトリを格納し、キャッシュメモリ部14にアクセスする前に必ず、共有メモリ部15に格納したディレクトリにアクセス中を示すビットを立てる。これにより、キャッシュメモリ部14内の同じアドレスには同時に1つのアクセス要求しか発行されないため、二重に書き込むデータの一方のデータの書き込み処理を行っている間に、もう一方のデータが他のアクセスバスからの書き込み要求によって書き換えられるのを防ぐことができる。

【0086】[実施例3] ディスクアレイ制御装置1では、キャッシュメモリ部14を複数設けた場合、あるキャッシュメモリ部14から別のキャッシュメモリ部14へデータをコピーする機能が要求される。この機能は、以下に述べる手順で実現できる。

【0087】アクセス開始の指示を受けたホストIF102あるいはドライブIF103は、制御線1：211によりCMアクセス制御部104内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとと



もに、データ線210を通して2つのアドレス、コマンドを送出する。2つのうちの1つのアドレス及びコマンドは、コピー元のアドレスとリードコマンドで、もう1つのアドレス及びコマンドはコピー先のアドレスとライトコマンドである。ここでは、CMコントローラa:107をコピー元、CMコントローラbをコピー先として説明する。

【0088】CMアクセス制御部104は、データ線210を通して送られてきた2つのアドレス、コマンドをバッファ303に格納する。データ転送制御部310はアービトレーションを行ってバス1F301の使用権を決定し、セクタ302を切り替える。

【0089】図12は、2つのキャッシュメモリ部間でデータをコピーする場合の、CMアクセス制御部104からCMコントローラa、b:107へのアクセスの流れを示している。CMアクセス制御部104内のデータ転送制御部310は、アービトレーションによってアクセスバス0:135の使用権が決定されると、制御線2:212によってセクタ部13内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ801)。続いて、アドレス及びコマンドを2つ連続して送出する(ステップ802)。

【0090】セクタ部13内のデータ転送制御部315は、CMアクセス制御部104からREQ信号を受け取ると、次にアクセスバス0:135を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行う(ステップ803)。アービトレーションの結果、CMコントローラa、b:107への2つのアクセスバス1:136への接続権の両方を得たら、データ転送制御部315はセクタ306を切り替える(ステップ804)とともに、制御線2:212により、CMアクセス制御部104内のデータ転送制御部310へ、CMコントローラa、b:107両方への接続権が得られたことを示す信号(ACK)を返す(ステップ805)。次にデータ転送制御部315は、制御線3:213によってCMコントローラa、b:107内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ806a、806b)。続いて、CMコントローラa、b:107へそれぞれのアドレス及びコマンドを送出する(ステップ807a、807b)。

【0091】CMアクセス制御部104はACK信号を受けると、アクセスの終了を知らせるステータスの受信待ち状態に入る。

【0092】コピー元のCMコントローラa:107内のデータ転送制御部315は、制御線3:213によってREQ信号を受け取ると、次にアクセスバス1:136を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要

求に基づいてアービトレーションを行い(ステップ808)、セクタ304を切り替える。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ809)。次に、メモリモジュール109からデータを読み出し(ステップ810)、セクタ304を介してアクセスバス1:136へ送出する(ステップ811a)。

【0093】セクタ部13は、アクセスバス1:136を通してCMコントローラa:107から送られてきたデータをCMコントローラbに繋がるアクセスバス1:136へ送出する。(ステップ811b)。

【0094】CMコントローラb:107内のデータ転送制御部315は、制御線3:213によってREQ信号を受け取ると、次にアクセスバス1:136を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析する。その後、データを受信し始めるのを待って、メモリアccessのアービトレーションに参加する(ステップ812)。アクセスバス1:136を通して送られてくるデータはバッファ303に格納する。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ813)。次に、バッファ303からデータを読み出し、セクタ304を介してメモリモジュール109へ書き込む(ステップ814)。

【0095】CMコントローラa、b:107は、それぞれCMメモリモジュール109へのアクセスが終了すると、メモリアccessの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ815、818)。次に、ステータスをセクタ部13へ送出する(ステップ816、819)。

【0096】セクタ部13内のデータ転送制御部315はステータスを受け取ると、CMコントローラa、b:107へのREQ信号をそれぞれオフする(ステップ817、821)。また、セクタ部13は2つのCMコントローラa、b:107の両方からステータスを受け取ったら、それらを続けてCMアクセス制御部へ送出する(ステップ820)。CMアクセス制御部104内のデータ転送制御部310は2つのステータスを受け取ると、セクタ部13へのREQ信号をオフする(ステップ822)。セクタ部13内のデータ転送制御部315はCMアクセス制御部104からのREQ信号のオフを確認すると、CMアクセス制御部104へのACK信号をオフする(ステップ823)。

【0097】CMアクセス制御部104内のデータ転送制御部310はステータスを受け取ると、制御線1:2



11により、ホストIF102あるいはドライブIF103へキャッシュメモリ部14へのアクセスの終了を報告する。

【0098】ディスクアレイ制御装置1ではまた、1つのキャッシュメモリ部14内のあるアドレスから別のアドレスへデータをコピーする機能も要求される。

【0099】この機能は、図9で示したデータの書き込み時の手順において、ステップ511のメモリモジュール109へのライトアクセスの代わりに、メモリモジュール109からデータを読み出してCMコントローラ107内のバッファ303に格納し、続けてそのデータをメモリモジュール109へ書き込むという処理を行うことによって実現できる。

【0100】〔実施例4〕図1において、2つの共有メモリ部15間で、メモリ領域の全部または、一部を二重化し、共有メモリ部15へのデータの書き込み時に二重化した2つの領域に同じデータを書き込むことにより、データの信頼性を上げることが可能となる。

【0101】二重化した2つの共有メモリ部15へデータを書き込む場合の手順は、以下のようになる。

【0102】マイクロプロセッサ101は、制御線5:221によりSMアクセス制御部105内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとともに、データ線220を通して2つのアドレス、コマンドと、1つのデータを送出する。

【0103】SMアクセス制御部105は、データ線220を通して送られてきた2つのアドレス、コマンドと、1つのデータをバッファ303に格納する。データ転送制御部310はアービトレーションを行ってバスIF301の使用権を決定し、セクタ302を切り替える。

【0104】図13は、二重化した2つの共有メモリ部15へデータを書き込む場合の、SMアクセス制御部105から2つのSMコントローラ108へのアクセスの流れを示している。2つの共有メモリ部を二重化する場合、一方をマスタ、もう一方をスレーブに設定する。SMアクセス制御部105内のデータ転送制御部310は、アービトレーションによって、まずマスタ側のSMコントローラ108へのアクセスバス2:137の使用権を決定し、制御線6:222によってマスタ側のSMコントローラ108内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ901)。続いて、アドレス、コマンド、及びデータを連続して送付する(ステップ902)。

【0105】マスタ側のSMコントローラ108内のデータ転送制御部315は、制御線6:222によってREQ信号を受け取ると、次にアクセスバス2:137を通して送られてくるアドレス、コマンド、及びデータを受信し、addr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行い(ステップ

903)、セクタ309を切り替える。データはバッファ303に格納する。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ904)。次に、バッファ303からデータを読み出し、セクタ309を介してメモリモジュール109へ書き込む(ステップ905)。

【0106】メモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ906)。次に、ステータスをSMアクセス制御部105へ送付する(ステップ907)。

【0107】SMアクセス制御部105はステータスを受け取ったら、マスタ側のSMコントローラ108へのアクセスバス2:137の使用権を開放せずに、アービトレーションによって、スレーブ側のSMコントローラ108へのアクセスバス2:137の使用権を決定する。その後のスレーブ側のSMコントローラ108へのアクセス手順(ステップ908~914)は、マスタ側のSMコントローラ108へのアクセス手順(ステップ901~907)と同様である。

【0108】SMアクセス制御部105内のデータ転送制御部310はスレーブ側のSMコントローラ108からステータスを受け取ると、マスタ側とスレーブ側両方のSMコントローラ108へのREQ信号をオフする(ステップ715a、715b)。

【0109】二重化した共有メモリ部15にデータを二重に書き込む場合、上記のように共有メモリ部15をマスタとスレーブに分け、マスタ側、スレーブ側の順でデータを書き込み、スレーブ側へのデータの書き込みが終わるまで両方のアクセスバス2:137を開放しない。これによりデータを書き込む順番が保証され、二重に書き込むデータの一方のデータの書き込み処理を行っている間に、もう一方のデータが他のアクセスバスからの書き込み要求によって書き換えられるのを防ぐことができる。

【0110】SMアクセス制御部105内のデータ転送制御部310はステータスを受け取ると、制御線5:221により、マイクロプロセッサ101へ共有メモリ部15へのアクセスの終了を報告する。

【0111】〔実施例5〕図14に、実施例1のディスクアレイ制御装置1のチャンネルIF部11、ディスクIF部12、セクタ部13、キャッシュメモリ部14、共有メモリ部15を実装するときの構成を示す。

【0112】チャンネルIF部11、ディスクIF部12、セクタ部13、キャッシュメモリ部14、共有メモリ部15は、それぞれ独立したパッケージ(PK)、すなわち、チャンネルIFPK1、ディスクIFPK2、

セクタPK3、メモリPK4にそれぞれ実装する。異なるキャッシュメモリ部14は、異なるパッケージ上に実装する。また、異なる共有メモリ部15も、異なるパッケージ上に実装する。1つのキャッシュメモリ部14と1つの共有メモリ部15は同一のパッケージ上に実装しても問題ない。図14では、1つのキャッシュメモリ部14と1つの共有メモリ部15を同一のメモリPK4上に実装した例を示している。

【0113】チャンネルIFPK1、ディスクIFPK2、セクタPK3、メモリPK4は、ブラッタ5上に実装し、それを筐体に搭載する。

【0114】ここで、各PKをブラッタ5に実装するときには、セクタPK3を中心付近より外側、好ましくはブラッタの両端に配置することが重要となる。このような配置により、各PK間を結ぶ線をブラッタ5上に配線する際、ブラッタ5全体にわたって配線の密度を均一にでき、ブラッタ上の配線を容易になる。

【0115】本実施例では、セクタ部13をセクタPK3に実装して、ブラッタ5の両端に配置するとした。しかし、セクタ部13をパッケージに実装せず、ブラッタの両端に直接実装しても問題ない。

【0116】〔実施例6〕図1に示す実施例1のディスクアレイ制御装置1において、SMアクセス制御部105とSMコントローラ108間をアクセスバス2:137で接続する代わりに、図7に示すように、2本の共有バス130を介して接続する。SMアクセス制御部105からは、2本の共有バス130それぞれに1本ずつ接続バスを設ける。また、SMコントローラ108からも、2本の共有バス130それぞれに1本ずつ接続バスを設ける。SMアクセス制御部105及びSMコントローラ108は、それぞれ2つのアービタを有している。2つのアービタはそれぞれ2本の共有バスのアービトレーション用のアービタである。SMアクセス制御部105からSMコントローラ108へのアクセスの際は、複数のアービタの内の1つがマスタとなり、共有バス130の使用権のアービトレーションを行う。そして、使用権を得たSMアクセス制御部105がSMコントローラ108にアクセスを行う。また、アービタをSMアクセス制御部105及びSMコントローラ108内に設ける代わりに、独立した回路として共有バス130に直接接続しても問題ない。

【0117】共有バス接続では、バスのデータ幅を広げることによりデータ転送速度を上げることが可能であり、共有メモリ部15へのアクセス時間を短縮可能である。

【0118】上述したように、図1におけるアクセスバス2:137のデータ幅は、アクセスバス0:135のデータ幅よりも2倍以上小さくすることができるので、図1のようにチャンネルIF部11及びディスクIF部12と共有メモリ部15との間をスター接続(1対1接

続)しても、共有メモリ部を実装するLSIのピンネックの問題が生じるケースは少ない。とはいえ、アクセスバス2:137の本数が増え過ぎてアクセスバス2:137を実装できないという問題が生じる可能性もある。そうした場合、本実施例の共有バス接続が有効となる。

【0119】なお、本実施例では、耐障害性の向上という観点から、セクタ部13、キャッシュメモリ部14、及び共有メモリ部15をそれぞれ二重化しているが、これらを二重化しなくても、上述の効果をえられることは言うまでもない。

【0120】

【発明の効果】本発明によれば、チャンネルIF部、ディスクIF部-キャッシュメモリ間のアクセスバスについてはスループットを高くできる。また、チャンネルIF部、ディスクIF部-共有メモリ間のアクセスバスについてはスループットを高くでき、かつアクセス時間を短くできる。これによって、スループットが高く、かつ応答時間の短いディスクアレイ制御装置を提供することが可能となる。

【図面の簡単な説明】

【図1】本発明によるディスクアレイ制御装置の構成を示す図。

【図2】従来のディスクアレイ制御装置の構成を示す図。

【図3】従来のディスクアレイ制御装置の他の構成を示す図。

【図4】本発明によるディスクアレイ制御装置内のCMアクセス制御部の構成を示す図。

【図5】本発明によるディスクアレイ制御装置内のSMアクセス制御部の構成を示す図。

【図6】本発明によるディスクアレイ制御装置内のセクタ部の構成を示す図。

【図7】本発明によるディスクアレイ制御装置内のキャッシュメモリ部の構成を示す図。

【図8】本発明によるディスクアレイ制御装置内の共有メモリ部の構成を示す図。

【図9】キャッシュメモリ部へのデータの書き込み時の手順を示す図。

【図10】共有メモリ部へのデータの書き込み時の手順を示す図。

【図11】二重化した2つのキャッシュメモリ部へデータを二重に書き込む時の手順を示す図。

【図12】1つのキャッシュメモリ部から別のキャッシュメモリ部へデータをコピーする時の手順を示す図。

【図13】二重化した2つの共有メモリ部へデータを二重に書き込む時の手順を示す図。

【図14】本発明によるディスクアレイ制御装置内の実装の構成を示す図。

【図15】本発明によるディスクアレイ制御装置内の他の構成を示す図。

【図16】本発明によるディスクアレイ制御装置内の他の構成を示す図。

【図17】本発明によるディスクアレイ制御装置の構成を示す図。

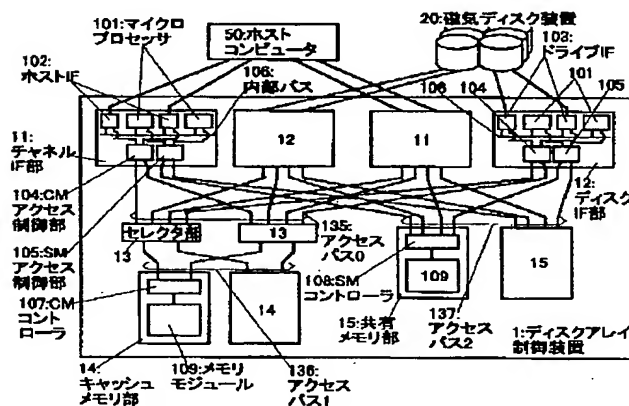
【符号の説明】

1…ディスクアレイ制御装置、11…チャネルIF部、12…ディスクIF部、13…セクタ部、14…キャッシュメモリ部、15…共有メモリ部、20…磁気ディ

スク装置、50…ホストコンピュータ、101…マイクロプロセッサ、102…ホストIF、103…ドライブIF、104…CMアクセス制御部、105…SMアクセス制御部、106…内部バス、107…CMコントローラ、108…SMコントローラ、109…メモリモジュール、135…アクセスバス0、136…アクセスバス1、137…アクセスバス2。

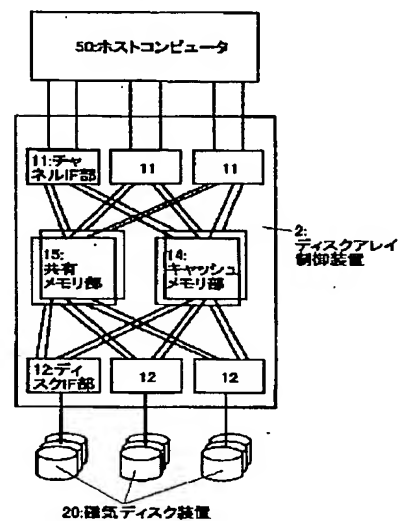
【図1】

図1



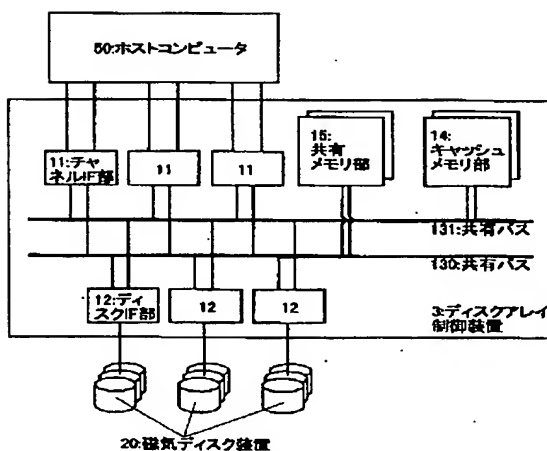
【図2】

図2



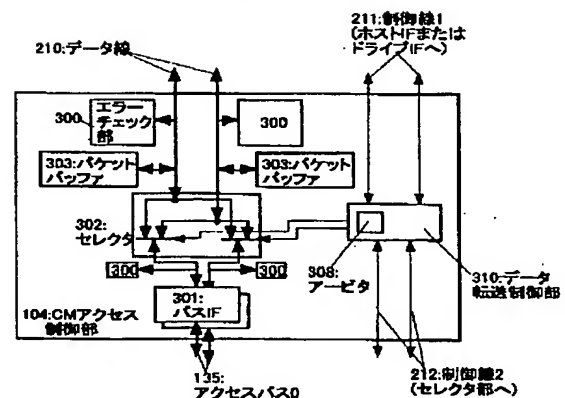
【図3】

図3



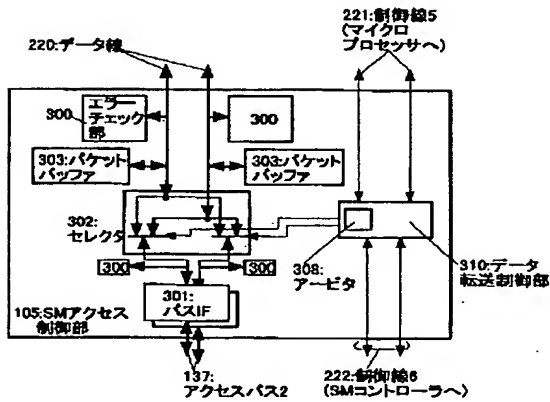
【図4】

図4



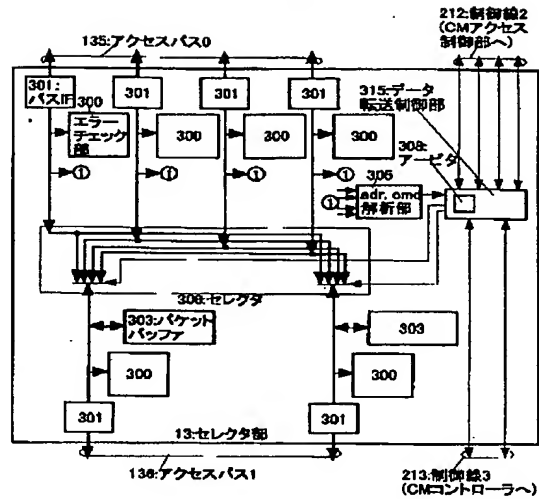
【図5】

図5



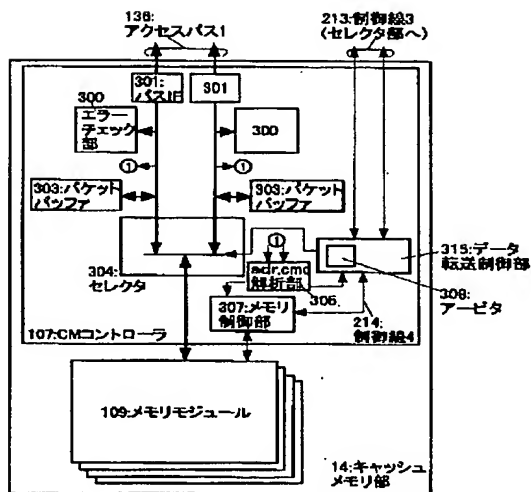
【図6】

図6



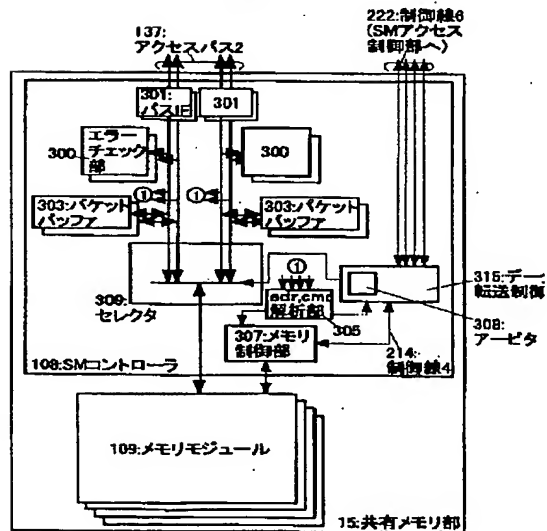
【図7】

図7



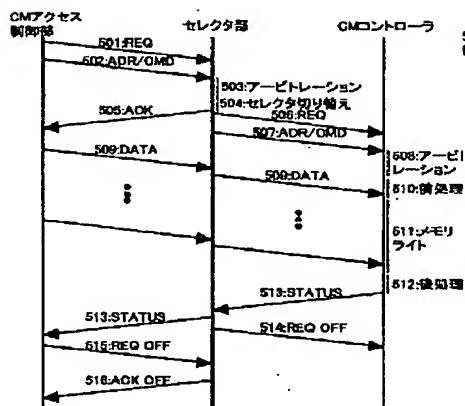
【図8】

図8



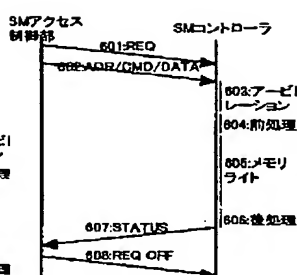
【図9】

図9



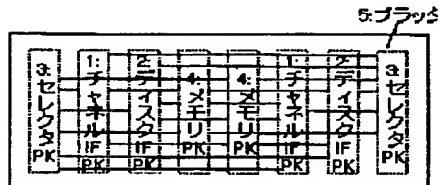
【図10】

図10



【図14】

図14

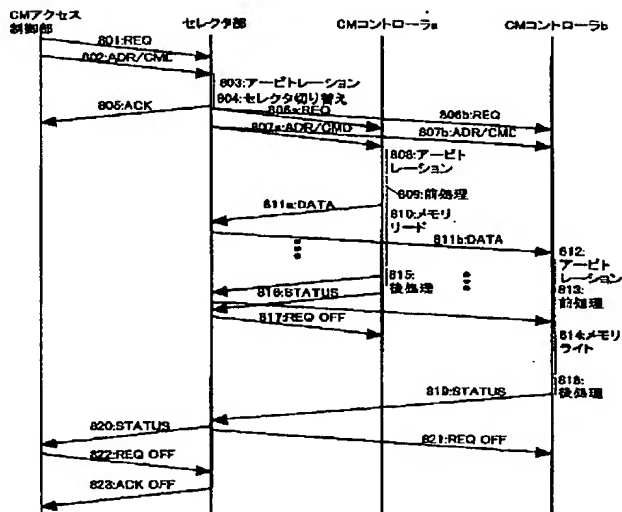
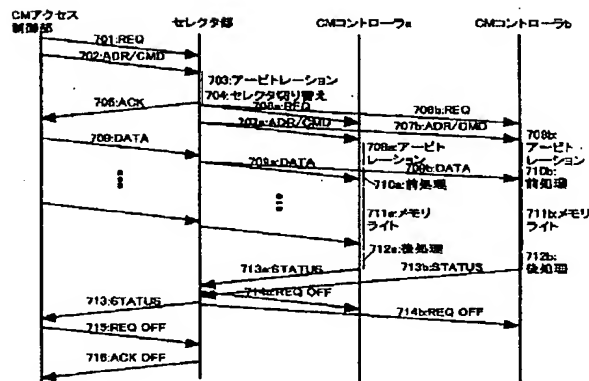


【図12】

図12

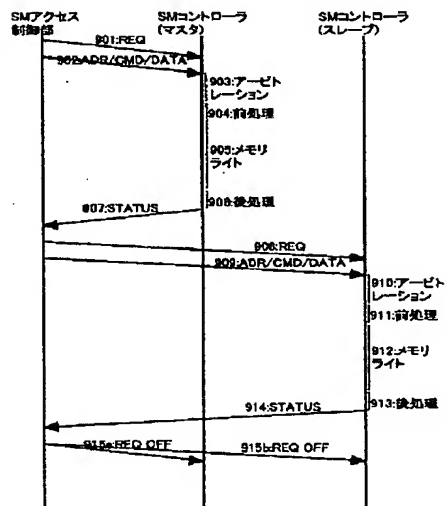
【図11】

図11



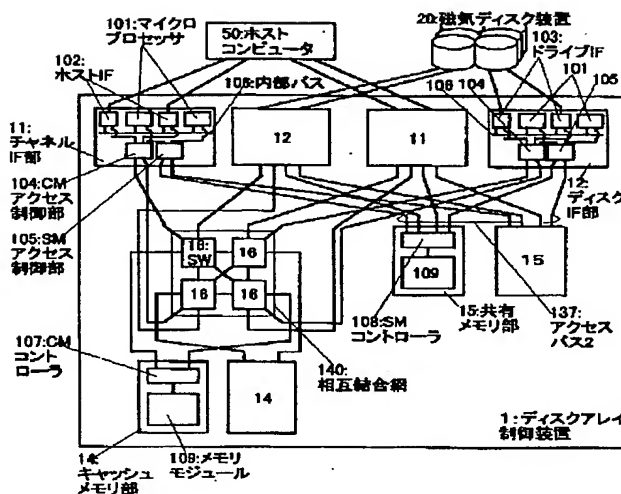
【図13】

図13



【図15】

図15



【図17】

【図16】

図16

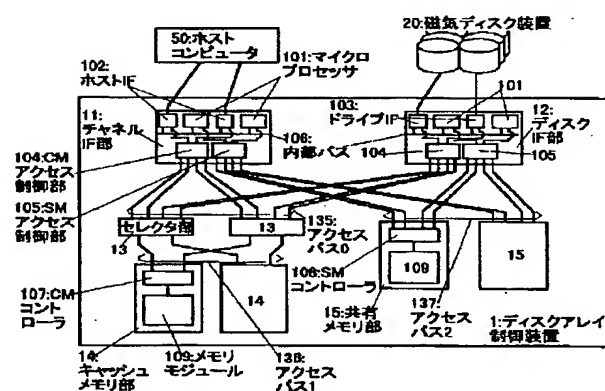
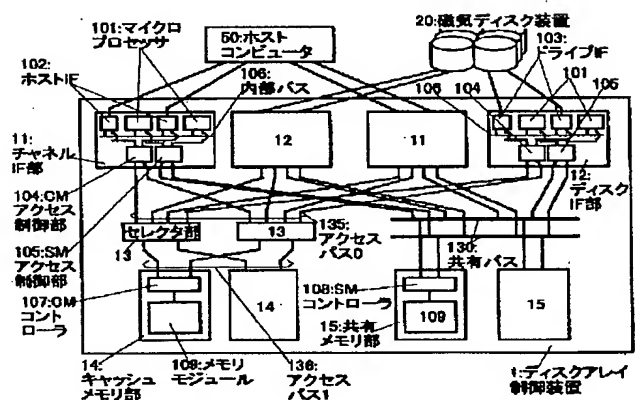


図17



フロントページの続き

(72)発明者 藤林 昭  
東京都国分寺市東恋ヶ窪一丁目280番地  
株式会社日立製作所中央研究所内

(72)発明者 金井 宏樹  
東京都国分寺市東恋ヶ窪一丁目280番地  
株式会社日立製作所中央研究所内

# ディスクアレイ制御装置

特開2000-99281

(72)発明者 箕輪 信幸

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

Fターム(参考) 5B065 BA01 CA12 CA30 CE11 CH01

W 1149

**COMPUTER SYSTEM AND ITS MAINTENANCE METHOD**

Patent Number: JP2002014878  
Publication date: 2002-01-18  
Inventor(s): ONO TETSURO  
Applicant(s): TOSHIBA CORP  
Requested Patent: ☐ JP2002014878  
Application Number: JP20000197802 20000630  
Priority Number(s):  
IPC Classification: G06F13/00; G06F1/24; G06F13/36; G06F15/16; G06F15/177  
EC Classification:  
Equivalents:

---

**Abstract**

---

**PROBLEM TO BE SOLVED:** To obtain necessary failure information even when the failure of a main processor is caused by making it possible to initialize a bus bridge device from a device other than a main processor.

**SOLUTION:** A server control bus (SMB) 4 is arranged between an I/O processor 173 and host bridges 13 and 14, and the environment setting processing of the host bridges 13 and 14 by the I/O processor 173 is operated through a server control bus (SMB) 4. The server control bus (SMB) 4 is constituted as an exclusive serial bus to be used for the maintenance control of this computer system so as to be made independent of a host bus 1 and PCI buses 2 and 3. Therefore, the server control bus 4 is used so that the host bridges 13 and 14 can be initialized under the control of the I/O processor 173 even when the failure of the main processor 11 is generated.

---

Data supplied from the esp@cenet database - I2



W1149

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-14878

(P2002-14878A)

(43) 公開日 平成14年1月18日 (2002.1.18)

| (51) Int.Cl.  | 識別記号  | F I           | テ-マコード* (参考)      |
|---------------|-------|---------------|-------------------|
| G 0 6 F 13/00 | 3 0 1 | G 0 6 F 13/00 | 3 0 1 N 5 B 0 4 5 |
| 1/24          |       | 13/36         | 3 1 0 E 5 B 0 5 4 |
| 13/36         | 3 1 0 | 15/16         | 6 2 0 A 5 B 0 6 1 |
| 15/16         | 6 2 0 | 15/177        | 6 7 8 B 5 B 0 8 3 |
| 15/177        | 6 7 8 | 1/00          | 3 5 0 A           |

審査請求 未請求 請求項の数8 O L (全 10 頁)

(21) 出願番号 特願2000-197802(P2000-197802)

(22) 出願日 平成12年6月30日 (2000.6.30)

(71) 出願人 000003078

株式会社東芝

東京都港区芝浦一丁目1番1号

(72) 発明者 大野 哲朗

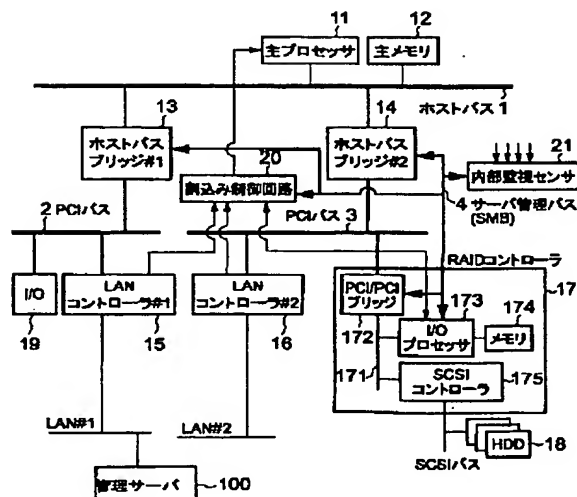
東京都青梅市末広町2丁目9番地 株式会

社東芝青梅工場内

# (57) 【要約】

【課題】主プロセッサ以外のデバイスからバスブリッジ装置の初期化処理を行えるようにし、主プロセッサの障害発生時にも必要な障害情報を取得できるようにする。

【解決手段】I/Oプロセッサ173とホストブリッジ13、14との間にはサーバ管理バス (SMB) 4が配設されており、I/Oプロセッサ173によるホストブリッジ13、14の環境設定処理はサーバ管理バス (SMB) 4を介して行われる。サーバ管理バス (SMB) 4は本計算機システムの保守管理に使用される専用のシリアルバスであり、ホストバス1およびPCIバス2、3とは独立している。よって、このサーバ管理バス4を使用することにより、主プロセッサ11の障害発生時にも、I/Oプロセッサ173の制御によってホストブリッジ13、14を初期化することができる。



【特許請求の範囲】

【請求項 1】 主プロセッサと、各種デバイスが接続されるバス間を相互接続するバスブリッジ装置とを有してなる計算機システムにおいて、

前記バスブリッジ装置に接続され、前記複数のバスとは独立して動作可能なシステム管理用バスと、  
前記主プロセッサの動作停止を伴う前記計算機システムの障害発生時に、バスを介した前記各種デバイス間の通信が可能になるように前記バスブリッジ装置の動作環境設定に必要な環境設定情報を前記システム管理用バスを介して前記バスブリッジ装置に設定する障害監視手段を具備することを特徴とする計算機システム。

【請求項 2】 前記障害監視手段は、  
前記計算機システムの稼働中に前記バスブリッジ装置からその環境設定情報を予め取得する手段を含み、  
前記障害発生時に、前記取得した環境設定情報を前記システム管理用バスを介して前記バスブリッジ装置に設定することを特徴とする請求項 1 記載の計算機システム。

【請求項 3】 主プロセッサと、各種デバイスが接続されるバス間を相互接続するバスブリッジ装置とを有してなる計算機システムにおいて、  
前記バスブリッジ装置に接続され、前記複数のバスとは独立して動作可能なシステム管理用バスと、  
前記計算機システムの I/O モジュールとして機能する I/O デバイス内に設けられ、前記主プロセッサの動作停止を伴う前記計算機システムの障害発生時に、前記バスブリッジ装置の動作環境設定に必要な環境設定情報を前記システム管理用バスを介して前記バスブリッジ装置に設定する I/O プロセッサとを具備することを特徴とする計算機システム。

【請求項 4】 前記 I/O プロセッサは、  
前記計算機システムの稼働中に前記バスブリッジ装置からその環境設定情報を予め取得する手段を含み、  
前記障害発生時に、前記予め取得した環境設定情報を前記システム管理用バスを介して前記バスブリッジ装置に設定することを特徴とする請求項 3 記載の計算機システム。

【請求項 5】 前記各種デバイスからの割り込み要求に応じて前記主プロセッサに割り込み信号を発行する割り込み制御手段をさらに具備し、  
前記 I/O プロセッサは、前記障害発生時に前記システム管理用バスを介して前記割り込み制御手段を制御して、前記割り込み信号の発行先を前記主プロセッサから前記 I/O プロセッサに切り替える手段をさらに含むことを特徴とする請求項 3 記載の計算機システム。

【請求項 6】 前記 I/O プロセッサは、  
前記バスブリッジ装置の環境設定処理を実行した後、前記バスブリッジ装置を介したバスサイクルによって前記各種デバイスから前記計算機システムの保守管理に必要な障害情報を取得する手段と、

前記取得した情報を前記複数のバスのいずれかに接続された通信デバイスを経由して外部に通知する手段とをさらに具備することを特徴とする請求項 3 記載の計算機システム。

05 【請求項 7】 主プロセッサと、各種デバイスが接続されるバス間を相互接続するバスブリッジ装置とを有してなる計算機システムの保守管理方法であって、  
前記主プロセッサの動作停止を伴う前記計算機システムの障害発生を検出するステップと、

10 前記障害発生が検出されたとき、前記バスを介した前記各種デバイス間の通信が可能になるように、前記バスブリッジ装置の動作環境設定に必要な環境設定情報を前記複数のバスとは独立して動作可能なシステム管理用バスを介して前記バスブリッジ装置に設定するステップとを具備することを特徴とする保守管理方法。

【請求項 8】 前記計算機システムの稼働中に前記バスブリッジ装置からその環境設定情報を予め取得するステップをさらに具備し、  
前記障害発生を検出時には、前記取得した環境設定情報を前記システム管理用バスを介して前記バスブリッジ装置に設定することを特徴とする請求項 7 記載の保守管理方法。

【発明の詳細な説明】

【0001】

25 【発明の属する技術分野】本発明は計算機システムおよびその保守管理方法に関し、特にバスブリッジ装置を有する計算機システムおよびその保守管理方法に関する。

【0002】

【従来の技術】従来より、計算機システムの保守管理技術としてはサービスプロセッサ (SVP) を用いたものが知られている。サービスプロセッサは保守機能 (障害通知、情報取得、障害修復など) を実現するための専用の監視装置であり、計算機システム本体とは独立した専用ハードウェアによって実現されている。このため、計算機システム本体の障害を外部に確実に通知することはできるものの、その反面コストの増大を招くという問題がある。

【0003】ところで、最近の計算機システムにおいては、階層構造化された複数のバスが設けられているのが通常である。これら複数のバス間はバスブリッジ装置によって接続されており、互いに異なるバス上に接続されたデバイス間の通信は全てバスブリッジ装置を介して行われる。また、バスブリッジ装置にはバスアービタなどのバス制御に関する機能が内蔵されている場合が多く、  
45 この場合には同一バス上のデバイス同士であってもその間の通信には通常はバスブリッジ装置が必要とされる。

【0004】バスブリッジ装置の環境設定処理 (初期化処理) は主プロセッサ (CPU) によって行われる。このため、もし主プロセッサの動作が停止されるような障害が発生すると (主プロセッサ自体の障害、主プロセッ

サが接続されたホストバスの障害)、バスブリッジ装置の動作環境を正しく設定することができなくなる。この場合、ハードウェア的には問題の無い他の各種I/Oデバイスがバス上に存在する場合であっても、それらデバイスへのアクセスが出来なくなり、保守管理に有効な情報を取得できなくなるという問題が生じる。

【0005】

【発明が解決しようとする課題】そこで、最近では、バスブリッジ装置そのものの機能を用いずに、そのバスブリッジ装置の上位側バス/下位側バスを外部から直接アクセス制御するための仕組みが提案されている(特開平11-259383号公報)。

【0006】この技術では、バスブリッジ装置のバス制御機能自体は用いていないので、バスブリッジ装置自体が正常動作せずとも必要な情報をバス上のデバイスから取得できる。しかし、計算機システム本体とは独立した専用ハードウェア(サービプロセッサ)を用意し、その専用ハードウェアがバスブリッジ装置に代わってバス制御およびI/O制御を全て行うという大がかりな構成が必要となるので、多大なコスト増を招くことになる。

【0007】本発明は上述の事情に鑑みてなされたものであり、主プロセッサ以外のデバイスからもバスブリッジ装置の初期化処理を行えるようにし、バスブリッジ装置自体の機能を用いて各種デバイスに対するアクセスを実現可能な計算機システムおよびその保守管理方法を提供することを目的とする。

【0008】

【課題を解決するための手段】上述の課題を解決するため、本発明は、主プロセッサと、各種デバイスが接続されるバス間を相互接続するバスブリッジ装置とを有してなる計算機システムにおいて、前記バスブリッジ装置に接続され、前記複数のバスとは独立して動作可能なシステム管理用バスと、前記主プロセッサの動作停止を伴う前記計算機システムの障害発生時に、バスを介した前記各種デバイスに対するアクセスが可能になるように前記バスブリッジ装置の動作環境設定に必要な環境設定情報を前記システム管理用バスを介して前記バスブリッジ装置に設定する障害監視手段を具備することを特徴とする。

【0009】この計算機システムにおいては、複数のバスとは独立して動作可能なシステム管理用バスが設けられており、主プロセッサの障害発生時には、そのシステム管理用バスを介してバスブリッジ装置に対する環境設定情報の設定が行われ、これによってバスブリッジ装置が初期化される。主プロセッサの障害によりバスブリッジ装置を初期化できないような環境であっても、主プロセッサ以外のバス上の他のデバイスがハードウェア的に問題がない状態であれば、それらデバイスに対するアクセスをバスブリッジ装置自体のバス制御機能を利用して行うことが可能となる。よって、各種デバイスから保守

管理に必要な情報を取得したり、障害発生を通信用デバイスを通じて外部に通知するといった保守管理動作を容易に行うことができる。

【0010】障害発生時にバスブリッジ装置に設定する環境設定情報は、計算機システムが正常に動作している時に予めバスブリッジ装置から取得しておき、それを障害発生時に利用することが好ましい。これにより、正しい環境設定情報を容易にバスブリッジ装置に設定することが可能となる。

【0011】また、本発明は、主プロセッサと、各種デバイスが接続されるバス間を相互接続するバスブリッジ装置とを有してなる計算機システムにおいて、前記バスブリッジ装置に接続され、前記複数のバスとは独立して動作可能なシステム管理用バスと、前記計算機システムのI/Oモジュールとして機能するI/Oデバイス内に設けられ、前記主プロセッサの動作停止を伴う前記計算機システムの障害発生時に、前記バスブリッジ装置の動作環境設定に必要な環境設定情報を前記システム管理用バスを介して前記バスブリッジ装置に設定するI/Oプロセッサとを具備することを特徴とする。

【0012】この計算機システムにおいては、その計算機システム内の通常のI/Oデバイス内に設けられたI/Oプロセッサにバスブリッジ装置に対する環境設定機能が搭載されており、主プロセッサの動作停止を伴う障害発生が発生すると、I/Oプロセッサ側からの制御でバスブリッジ装置にその動作環境設定に必要な環境設定情報が設定される。環境設定情報の設定は、上述したように、複数のバスとは独立して動作可能なシステム管理用バスを介して行われるので、通常バスサイクルが正常に実行できない環境下でもバスブリッジ装置を機能回復させることができる。よって、そのバスブリッジ装置の機能を利用してバス上の各種デバイスをアクセスすることができるので、計算機本体とは独立した専用ハードウェアを用いることなく、計算機システムに通常設けられているハードウェアモジュールのみにより容易に保守管理のための動作を実行することが可能となる。

【0013】また、前記I/Oプロセッサに、前記バスブリッジ装置の環境設定処理を実行した後、前記バスブリッジ装置を介したバスサイクルによって前記各種デバイスから前記計算機システムの保守に必要な障害情報を取得する手段と、前記取得した障害情報を前記複数のバスのいずれかに接続された通信デバイスを経由して外部に通知する手段とをさらに設けることにより、保守管理に必要なほとんど全ての動作をI/Oプロセッサによって実現することが可能となる。また、外部への情報通知にはバス上に接続された通信デバイスを用いているので、計算機本体とは独立した保守管理専用の通信デバイスを用意する必要もない。

【0014】また、割り込み信号の発行先を主プロセッサから前記I/Oプロセッサに切り替える機構をさらに

容易することにより、各デバイスからの割り込みをI/Oプロセッサ側で処理することが可能となり、計算機システム内のハードウェア動作の制御を主プロセッサに代わってI/Oデバイス側で容易に行うことが可能となる。

【0015】

【発明の実施の形態】以下、図面を参照して本発明の実施形態を説明する。図1には、本発明の一実施形態に係る計算機システムの構成が示されている。この計算機システムはサーバコンピュータとして使用されるものであり、図示のように、主プロセッサ(CPU)11、主メモリ12、第1および第2のホストバスブリッジ13、14、第1および第2のLANコントローラ15、16、RAIDコントローラ17、複数のHDDユニット18、他のI/Oデバイス19、割り込み制御回路20、および内部監視センサ21などから構成されている。

【0016】主プロセッサ(CPU)11は本計算機システム全体の動作を制御するためのものであり、主メモリ12上にロードされるオペレーティングシステムおよびアプリケーションプログラム等を実行する。ホストバスブリッジ13、14はそれぞれ主プロセッサ(CPU)11が接続されたホストバス1とPCIバス2、3間を相互接続するブリッジLSIであり、ホストバスとPCIバスと間でトランザクションを双方向で伝達する機能およびバスアービタなどを初めとするPCIバス制御機能を有している。PCIバス2、3は、各種I/Oデバイスを接続するためのI/Oバスとして用いられる。

【0017】ホストバスブリッジ13、14の各々はPCIデバイスであり、その動作環境設定に必要な環境設定情報(バスアドレス、PCIバス2、3上の各I/Oデバイスに割り当てられるI/Oアドレス空間、その他バス制御に必要な情報)は各々のホストバスブリッジ13、14内に設けられたコンフィグレーションレジスタに設定される。

【0018】ホストバスブリッジ13については、ホストバス1がプライマリーバスとなり、PCIバス2がセカンダリーバスとなる。ホストバス1上のバストランザクションで指定されるアドレスがPCIバス2上の各I/Oデバイスに対応するアドレス空間に属するとき、ホストバスブリッジ13は、ホストバス1上のバストランザクションをPCIバス2に伝達する。逆に、PCIバス2上のバストランザクションは、そのバストランザクションで指定されるアドレスがPCIバス2上の各I/Oデバイスに対応するアドレス空間に属さない時に、ホストバスブリッジ13によってホストバス1上に伝達される。

【0019】ホストバスブリッジ14については、ホストバス1がプライマリーバスとなり、PCIバス3がセ

カンダリーバスとなる。ホストバス1上のバストランザクションで指定されるアドレスがPCIバス3の各I/Oデバイスに対応するアドレス空間に属するとき、ホストバスブリッジ14は、ホストバス1上のバストランザクションをPCIバス3に伝達する。逆に、PCIバス3上のバストランザクションは、そのバストランザクションで指定されるアドレスがPCIバス3上の各I/Oデバイスに対応するアドレス空間に属さない時に、ホストバスブリッジ14によってホストバス1上に伝達される。

【0020】LANコントローラ15、16およびRAIDコントローラ17はI/Oデバイス19と同様に本計算機システムのI/Oモジュールとして設けられたPCIデバイスである。第1のLANコントローラ15はPCIバス2に接続されており、また第2のLANコントローラ16はPCIバス3に接続されている。これら第1および第2のLANコントローラ15、16はLAN接続のためのネットワークインターフェースを提供し、管理サーバ100との通信は、その管理サーバ100が存在するLANに接続された第1のLANコントローラ15を介して行われる。

【0021】RAIDコントローラ17は複数のHDDユニット18からなるディスクアレイの制御を例えばRAID5等の方式で実行するI/Oデバイスであり、図示のようにPCIバス3に接続されている。このRAIDコントローラ17は、内部PCIバス171、PCI/PCIブリッジ172、I/Oプロセッサ173、メモリ174、およびSCSIコントローラ175から構成されている。

【0022】PCI/PCIブリッジ172はPCIバス3と内部PCIバス171とを接続するブリッジである。内部PCIバス171にはI/Oプロセッサ173およびHDD制御のためのSCSIコントローラ175が接続されている。I/Oプロセッサ173はRAID制御に関する処理を行うためのプロセッサであり、メモリ174上のファームウェアに従って動作する。本実施形態においては、このI/Oプロセッサ173にホストブリッジ13、14を初めとするI/Oバス(PCIバス2、3)側に関する初期設定を行う機能、および障害通知、障害情報取得、障害修復などの障害保守機能が搭載されている。

【0023】I/Oプロセッサ173とホストブリッジ13、14との間には図示のようにサーバ管理バス(SMB)4が配設されており、I/Oプロセッサ173によるホストブリッジ13、14の環境設定処理(初期化処理)はサーバ管理バス(SMB)4を介して行われる。サーバ管理バス(SMB)4は本計算機システムの保守管理に使用される専用のシリアルバスであり、ホストバス1およびPCIバス2、3とは独立している。よって、このサーバ管理バス(SMB)4は、ホストブリ

ッジ13, 14のバス制御機能の状態およびホストバス1, PCIバス2, 3の状態に関係なく、独立して使用することができる。

【0024】サーバ管理バス(SMB)4には、さらに内部監視センサ21、割り込み制御回路20、およびRAIDコントローラ17のPCI/PCIブリッジ172も接続されており、I/Oプロセッサ173は、内部監視センサ21によって検知された本計算機システムの電源やファン等に関する情報をサーバ管理バス(SMB)4を介して取得したり、割り込み制御回路20およびPCI/PCIブリッジ172の動作をサーバ管理バス(SMB)4を介して制御することもできる。

【0025】割り込み制御回路20はLANコントローラ15, 16を初めとする各I/Oデバイスからの割り込み要求に応じて主プロセッサ11に割り込み信号を発行する。この割り込み制御回路20は、サーバ管理バス(SMB)4からの所定のコマンドに応じて、割り込み信号の発行先を主プロセッサ11からI/Oプロセッサ173に切り換え可能に構成されている。

【0026】なお、システム構成によっては、PCIバス2, 3にPCI/PCIブリッジを介して別のPCIバスがさらに接続される場合もある。また、実際には、ディスプレイコントローラやキーボードコントローラなど様々なI/OデバイスがPCIなどのI/Oバス上に接続されることになる。

【0027】本計算機システムにおいては、正常時には、主プロセッサ11がホストバスブリッジ13, 14、PCI/PCIブリッジ172、および各I/Oデバイスの初期設定(アドレス設定など)を行なう。また、各I/Oデバイスからの割り込み信号は、割り込み制御回路20経由で主プロセッサ11に伝えられ、必要なI/O処理を行なうことになる。

【0028】主プロセッサ11側で障害が発生した場合には、主プロセッサ11からI/Oプロセッサ173に対する定期的なアクセスがなくなり、I/Oプロセッサ173がタイムアウトを検出し、サーバ監視動作を開始する。I/Oプロセッサ173は、先ず、サーバ管理バス(SMB)4を経由して各バスブリッジ(本例では、ホストバスブリッジ13, 14、PCI/PCIブリッジ173)の初期化を行う。この初期化のための設定情報、つまり各バスブリッジのコンフィグレーションレジスタに設定すべきコンフィグレーション情報(環境設定情報)は、予めシステム正常動作時にRAIDコントローラ17のメモリ174上に格納しておく。さらに、I/Oプロセッサ173は、サーバ管理バス(SMB)4からの制御で割り込み制御回路20の割り込み信号出力先を切り換え、主プロセッサ11に出力していた割り込み信号をI/Oプロセッサ173側に切り替える。この処理により、I/Oプロセッサ173は、主プロセッサ11に代わって各I/Oデバイスからのイベントを割り

込み信号によって受け取ることが可能となる。また、ホストバスブリッジ13, 14も初期化しているので、PCIバスサイクルの実行により、PCIデバイス2, 3上の各I/Oデバイス、および主メモリ12へのアクセスも可能である。

【0029】こうした動作状態で、I/Oプロセッサ173は、主メモリ12上の情報や、サーバ管理バス(SMB)4上の内部監視センサ21による情報、さらには各I/Oデバイスの障害情報などをPCIバスサイクルにより、あるいはサーバ管理バス(SMB)4を介して取得し、それを障害情報としてLANコントローラ15経由でLAN上の管理サーバ100に通知することができる。

【0030】次に、図2を参照して、ホストバスブリッジ13, 14をサーバ管理バス(SMB)4側から初期設定するために必要なホストバスブリッジ13, 14の構成について説明する。

【0031】図2に示されているように、ホストバスブリッジ13, 14の各々には、プライマリPCIインタフェイス201、コンフィグレーションレジスタ202、サーバ管理バスインタフェイス(SMB I/F)203が設けられている。プライマリPCIインタフェイス201は、ホストバスブリッジ13, 14のプライマリバス側、つまりホストバス1との間のインタフェイス用の回路であり、コンフィグレーションレジスタ202に対するコンフィグレーション情報の設定は通常は主プロセッサ11によって実行されるコンフィグレーションサイクルによりホストバス1側から行われる。このコンフィグレーションサイクルにตอบสนองして、必要なコンフィグレーション情報がプライマリPCIインタフェイス201を介してコンフィグレーションレジスタ202に設定される。

【0032】本例においては、コンフィグレーションレジスタ202に対するアクセスは、サーバ管理バスインタフェイス(SMB I/F)203の働きによりサーバ管理バス(SMB)4側からも行うことができる。すなわち、I/Oプロセッサ173がサーバ管理バス(SMB)4を介して特定のコマンドをサーバ管理バスインタフェイス(SMB I/F)203に発行することにより、主プロセッサ11およびホストバス1を一切使用することなく、コンフィグレーションレジスタ202からの情報取得、およびコンフィグレーションレジスタ202への情報設定をI/Oプロセッサ173の制御によりサーバ管理バス(SMB)4側から行うことができる。

【0033】また、I/Oプロセッサ173の制御により、ホストバスブリッジ13, 14自体を必要に応じてサーバ管理バス(SMB)4側からリセットする事もできる。

【0034】次に、図3を参照して、割り込み制御回路

20の構成例について説明する。

【0035】割り込み制御回路20には、図示のように、割り込み制御部211、セクタ（SEL）212、およびサーバ管理バスインタフェイス（SMB I/F）213が設けられている。割り込み制御部211は、各I/Oデバイスからの割り込み要求信号を予め決められた割り込み優先順位に従って処理することによって一つの割り込み信号を発生する。セクタ（SEL）212は、割り込み制御部211からの割り込み信号の発行先を主プロセッサ11とI/Oプロセッサ173との間で切り替えるためのものであり、通常は、割り込み信号が主プロセッサ11に出力されるように構成されている。サーバ管理バス（SMB）4およびサーバ管理バスインタフェイス（SMB I/F）213を介してI/Oプロセッサ173からの所定のコマンドを受け取ると、セクタ（SEL）212によって割り込み信号の発行先が主プロセッサ11からI/Oプロセッサ173側に切り替えられる。

【0036】次に、図4を参照して、I/Oプロセッサ173の機能構成について説明する。以下の機能のほとんどはI/Oプロセッサ173によって実行されるファームウェアによって提供されるものである。

【0037】ウォッチドッグタイマ機能部311は主プロセッサ11からの定期的なアクセスが正常に行われているかを監視するためのものであり、所定期間経過してもアクセスが行われない時にはタイムアウト（障害発生検出）のイベントを発行する。このタイムアウトイベントは制御処理切換部312に送られ、制御部313によって行われるI/Oプロセッサ173の処理が通常時の処理からサーバ監視処理に切り替えられる。

【0038】制御部313には、図示のように、コンフィグレーション情報取得処理部314、RAID制御機能部315、およびサーバ監視処理部316が設けられている。コンフィグレーション情報取得処理部314およびRAID制御機能部315は通常処理時に実行されるものであり、コンフィグレーション情報取得処理部314はシステム稼働中に各バスブリッジのコンフィグレーション情報を取得してメモリ174に保存する処理を実行する。サーバ監視処理部316は、障害発生時に実行されるものであり、メモリ174上のコンフィグレーション情報を用いて各バスブリッジの初期化を行う初期化処理部401、割り込み先を切り替えるための割り込み制御回路切り替え処理部402、LANコントローラ15経由で障害発生を管理サーバ100に通知するための障害通知部403、および各I/Oや主メモリ12から障害情報を取得するための障害情報取得部404などから構成されている。

【0039】次に、図5および図6のフローチャートを参照して、通常時に行われるI/Oプロセッサ173の動作について説明する。

【0040】I/Oプロセッサ173は、サーバ管理バス（SMB）4を介してホストブリッジ13、14を初めとする各バスブリッジをアクセスすることにより、それら各バスブリッジのコンフィグレーションレジスタに設定されているコンフィグレーション情報を取得する

（ステップS101）。また、I/Oプロセッサ173は、必要に応じて、LANコントローラ15、16を初めとする各I/OデバイスをPCIバス経由でアクセスし、それらI/Oデバイスのコンフィグレーション情報も取得する（ステップS101）。次いで、I/Oプロセッサ173は、取得した各コンフィグレーション情報をメモリ174に保存する（ステップS103）。これらステップS101～S103の処理は、RAID制御機能を行う通常処理時における任意のタイミングで行われる。

【0041】また、通常処理時においては、図6のフローチャートに示すウォッチドッグ処理がバックグラウンドで実行される。

【0042】すなわち、図6のフローチャートに示されているように、I/Oプロセッサ173は、主プロセッサ11からの定期的なアクセスの有無を監視しており、予め決められた所定期間内に主プロセッサ11からの所定のアクセスがあるかを判定する（ステップS111）。主プロセッサ11による定期的なアクセスは、その主プロセッサ11が正常に動作していることを示すメッセージをI/Oプロセッサ173宛に送信するというものであり、これはオペレーティングシステムの一機能などとして実現されている。所定期間内に主プロセッサ11からのアクセスがあると、ウォッチドッグタイマ機能部311のタイマをクリアする処理が行われる（ステップS112）。

【0043】定期的なアクセスが途絶えると、ウォッチドッグタイマ機能部311はタイムアウトとなり（ステップS114のYES）、前述のサーバ監視処理機能が起動される。なお、ウォッチドッグタイマ機能部311は、I/Oプロセッサ173でなく、RAIDコントローラ17内の専用のハードウェアによって実現しても良い。

【0044】次に、図7のフローチャートを参照して、サーバ監視処理動作の具体的な手順を説明する。

【0045】サーバ監視処理においては、I/Oプロセッサ173は、まず、サーバ管理バス（SMB）4を通じてホストバスブリッジ13、14にそれぞれ対応するコンフィグレーション情報を設定して、それらホストバスブリッジ13、14の初期設定を行う（ステップS201）。次に、I/Oプロセッサ173は、サーバ管理バス（SMB）4を通じて自PCI/PCIブリッジ172に対応するコンフィグレーション情報を設定して、自PCI/PCIブリッジ172の初期設定を行う（ステップS202）。そして、I/Oプロセッサ173

は、サーバ管理バス（SMB）4を通じて割り込み制御回路20を制御することにより、割り込み信号の発行先を主プロセッサ11からI/Oプロセッサ173に切り替える（ステップS203）。その後、I/Oプロセッサ173は、第1のLANコントローラ15経由で管理サーバ100に対して障害発生を通知する（ステップS204）。この場合、I/Oプロセッサ173から管理サーバ100への障害発生通知は、自PCI/PCIブリッジ172、PCIバス3、ホストブリッジ14、ホストバス1、ホストブリッジ13、PCIバス2、およびLANコントローラ15という経路で実行される。LANコントローラ15経由での障害発生通知処理に際しては、LANコントローラ15からの割り込み信号がI/Oプロセッサ173側で処理されるので、I/Oプロセッサ173とLANコントローラ15との間の正常なデータ転送制御が可能となる。

【0046】管理サーバ100は、詳細な障害情報を取得するために、LANコントローラ15経由で障害情報取得要求メッセージをI/Oプロセッサ173に送出する。I/Oプロセッサ173は、LANコントローラ15から割り込み信号を受けたとき、LANコントローラ15をアクセスすることによって管理サーバ100からのメッセージを受信する。受信したメッセージが障害情報取得要求メッセージであるとき（ステップS205のYES）、I/Oプロセッサ173は、障害情報取得動作を開始する。

【0047】すなわち、I/Oプロセッサ173は、最初に、サーバ管理バス（SMB）4経由で内部監視センサ21から電源やファン等の状態に関する各種センサ情報を取得する（ステップS206）。次いで、I/Oプロセッサ173は、PCIバス3およびホストバス1経由で主メモリ12から各種ログなどのメモリ情報を取得する（ステップS207）。その後、I/Oプロセッサ173は、必要に応じて他の各種PCIデバイスの情報取得を行う（ステップS208）。そして、I/Oプロセッサ173は、取得した情報を障害情報としてLANコントローラ15経由で管理サーバ100に送信する（ステップS209）。

【0048】管理サーバ100は、障害情報を解析し、システム回復のために更に必要な処理をI/Oプロセッサ173に指示する。

【0049】以上のように、本実施形態によれば、I/Oプロセッサ173の制御によっての各バスブリッジを初期設定することができるので、各バスブリッジの機能を利用してバスサイクルを実行することが可能となり、障害情報の取得、通知などを容易に行うことが可能となる。

【0050】なお、本実施形態では、PCIバス2上のLANコントローラ15を通じて障害情報等の通知を行ったが、PCIバス3上のLANコントローラ16を通

じて管理サーバ100への通知を行う場合であっても、PCIバス3上のバスサイクルを正常に実行するため、あるいは主メモリ12からのメモリ情報の取得等のために、ホストバスブリッジ14の初期化は重要となる。

【0051】また、PCIバス2または3上に存在するディスプレイコントローラ経由で障害内容をディスプレイモニタに表示したり、あるいはモデムや他のデバイスを通じて外部に障害内容を通知することも可能である。さらに、障害復旧時にはLANやキーボードなどからの指示をI/Oプロセッサ173側で処理することにより、最適な処理を行うことが可能となる。

【0052】また、LANコントローラ15などのアクセス先のI/Oデバイスが動作しない場合には、それらから予め取得しておいたコンフィグレーション情報をI/Oプロセッサ173によって該当するI/Oデバイスに設定するようにしても良い。さらに、プロセッサを内蔵するI/Oデバイスであれば、RAIDコントローラに限らず、そのI/Oデバイス内のプロセッサを前述のI/Oプロセッサ173として利用することができる。

【0053】また、本発明は、上記実施形態に限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で種々に変形することが可能である。更に、上記実施形態には種々の段階の発明が含まれており、開示される複数の構成要件における適宜な組み合わせにより種々の発明が抽出され得る。例えば、実施形態に示される全構成要件から幾つかの構成要件が削除されても、発明が解決しようとする課題の欄で述べた課題が解決でき、発明の効果の欄で述べられている効果が得られる場合には、この構成要件が削除された構成が発明として抽出され得る。

【0054】

【発明の効果】以上説明したように、本発明によれば、主プロセッサ以外のデバイスからバスブリッジ装置の初期化処理を行うことにより、バスブリッジ装置自体の機能を用いて各種デバイスに対するアクセスを実現可能となる。よって、バス制御等の機能を併せ持つ専用のハードウェアを用意することなく、計算機システムの保守管理を容易に実現することができる。

【図面の簡単な説明】

【図1】本発明の一実施形態に係る計算機システムの構成を示すブロック図。

【図2】同実施形態の計算機システムに設けられているホストバスブリッジの機能構成を示すブロック図。

【図3】同実施形態の計算機システムに設けられている割り込み制御回路の構成を示すブロック図。

【図4】同実施形態の計算機システムに設けられているI/Oプロセッサの機能構成を示す図。

【図5】同実施形態の計算機システムに設けられているI/Oプロセッサによって通常処理時に行われるコンフィグレーション情報取得動作の手順を説明するためのフ



ローチャート。

【図6】同実施形態の計算機システムに設けられているI/Oプロセッサによって通常処理時に行われるウォッチドック処理動作の手順を説明するためのフローチャート。

【図7】同実施形態の計算機システムに設けられているI/Oプロセッサによって実行される障害監視処理動作の手順をするためのフローチャート。

【符号の説明】

1…ホストバス

2, 3…PCIバス

11…主プロセッサ

12…主メモリ

13, 14…ホストバスブリッジ

15, 16…LANコントローラ

05 17…LAI Dコントローラ

21…内部監視センサ

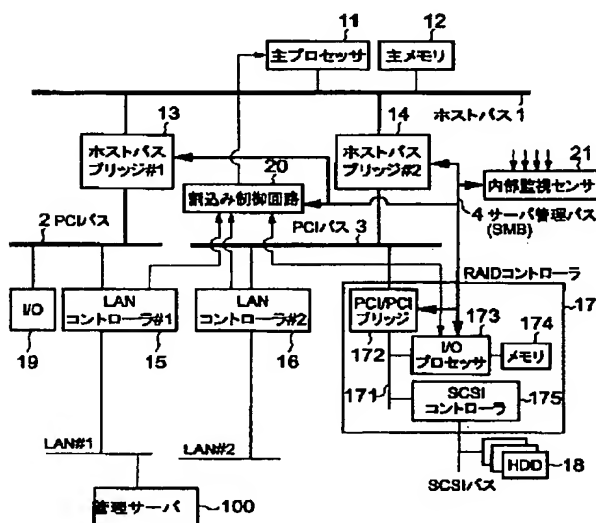
171…内部PCIバス

172…PCI/PCIブリッジ

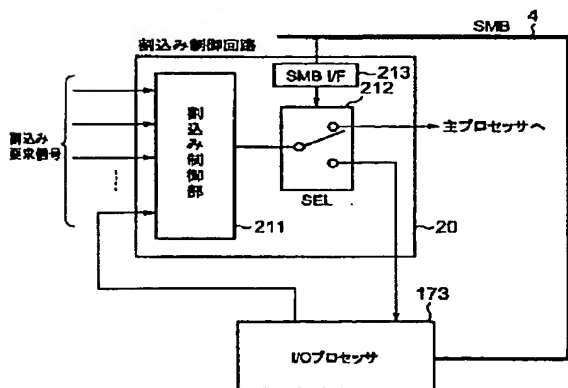
173…I/Oプロセッサ

10 175…SCSIコントローラ

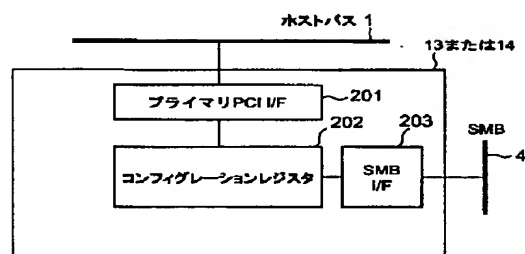
【図1】



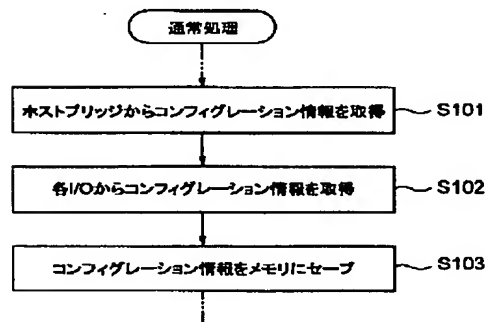
【図3】



【図2】

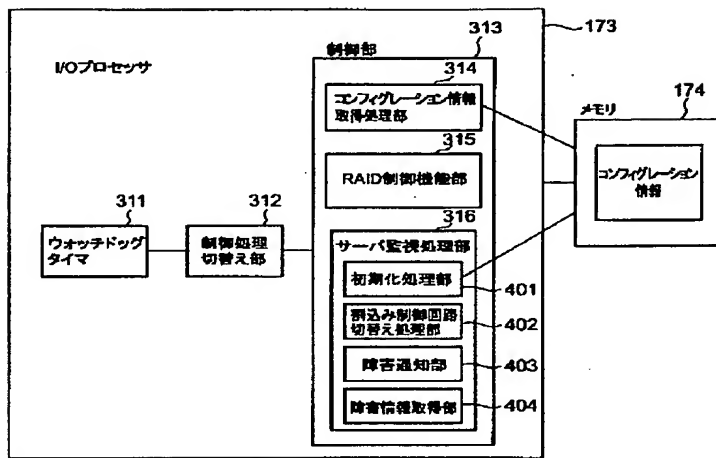


【図5】

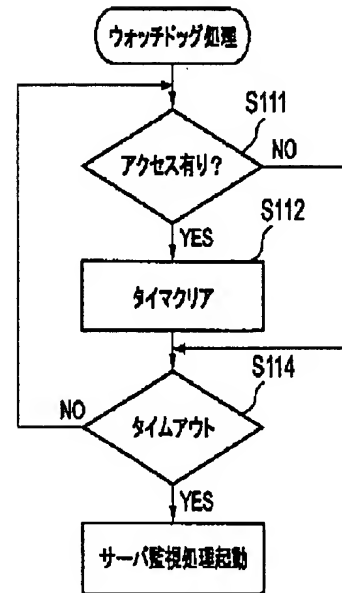




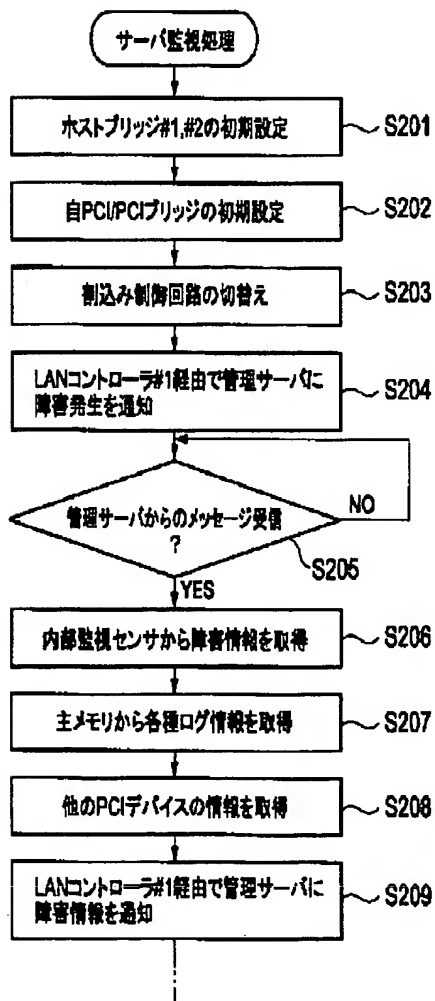
【図4】



【図6】



【図7】



フロントページの続き

F ターム(参考) 5B045 BB14 BB25 BB28 BB30 BB47  
FF03 HH06 JJ02 JJ08 JJ13  
JJ28 JJ46  
5B054 AA01 AA08 BB05 CC03  
5B061 FF01 GG02 QQ06  
5B083 AA08 BB01 BB03 CD07 CD10  
CE01 DD10 EE02

05